Matrix

2017年1月 第41期
微软亚洲研究院院刊

15 机器会产生自我意识吗？

08 培养一个人类医生需要八年，那么人工智能呢？

29 这里是中国人工智能的摇篮

Matrix

微软亚洲研究院院刊 2017年1月 第41期

黑科技

04 好玩的文本生成

08 培养一个人类医生至少需要八年，那么人工智能呢？

11 对偶学习：一种新的机器学习范式

13 如何在NLP领域第一次做成一件事

前沿求索

15 机器会产生自我意识吗？

18 17位微软科学家带你预测2017和2027

22 每一个人的人工智能

24 如何做好计算机视觉的研究？

文化故事

27 微软亚洲研究院研究员获选IEEE Fellow 和ACM Distinguished Member

媒体报道

29 这里是中国人工智能的摇篮，他们的故事仍在继续

挑战即机遇

不久前，比尔·盖茨接受了高晓松的一次视频采访，他在采访中多次提到人工智能技术目前的进展令人惊叹。盖茨表示，人工智能是一个前景广阔的领域，影响力惊人。然而，虽然人工智能在视觉和语音等感知能力方面已取得了长足的进步，但例如“阅读”等认知难题仍亟待解决。人工智能目前仍面临诸多挑战。

关于人工智能所面临的挑战，大家各持观点。我想与大家分享我近日读到的一篇文章《人工智能革命缺失的五个要素》，作者是机器智能企业CognitiveScale的董事长Manoj Saxena。他认为，尽管人工智能是当下热门话题，潜力巨大，但如果我们的眼界仅仅停留在当下的概念炒作，今后将面临诸多风险和问题。对此，Saxena提出应聚焦于五个关键领域，包括人工智能的应用、人才、测试、管理和体验。

- 在应用方面，受制于数据的复杂性和数据量的大小，数据科学家产生的商业价值往往无法达到业务部门预期，因此需要一套针对行业和流程的高价值使用案例，清晰地指出人工智能在哪些领域最能创造商业价值；
- 人才培养方面，除了算法设计能力，人工智能人才还需掌握数据科学、战略、设计、洞察力以及变化管理等多方面的能力；
- 质量保证是系统开发的重要一环，人工智能系统训练过程容易出现偏差，因此需要新的、带有原始参考标准的测试方法来保证系统的稳定运行；
- 人工智能系统应该是可管理并能清楚解释自身行为的，到目前为止，算法开发的动机都是提高性能，却牺牲了算法本身的可信度和可追溯性，我们被放进了一个不透明的“黑盒子”里，如果不加干预，这个问题只会变得更加严峻；
- 最后，系统的用户体验决定了它的价值，只有提供了令人愉悦的交互体验，人工智能系统才能获得用户的信任和使用。

诚然，人工智能目前面临诸多问题和挑战，但挑战即是机遇，机遇便是希望。这些问题需要产业界和学术界的同仁共同参与探索。我们也希望能有更多的年轻人加入我们，在前沿技术探索上更进一步，在普及人工智能的道路上继续向前迈进。

微软全球资深副总裁
微软亚洲研究院院长兼微软亚太研发集团主席

好玩的文本生成

作者： 微软亚洲研究院主管研究员 宋睿华

文本生成是比较学术的说法，通常在媒体上见到的“机器人写作”、“人工智能写作”、“自动对话生成”、“机器人写古诗”等，都属于文本生成的范畴。

二零一六年里，关于文本生成有许多的新闻事件，引起了学术界以外对这一话题的广泛关注。

2016年3月3日，MIT CSAIL[1]报道了，MIT计算机科学与人工智能实验室的一位博士后开发了一款推特机器人，叫DeepDrumpf，它可以模仿当时的美国总统候选人Donald Trump来发文。

2016年3月22日，日本共同社报道，由人工智能创作的小说作品《机器人写小说的那一天》入围了第三届星新一文学奖的初审。这一奖项以被誉为“日本微型小说之父”的科幻作家星新一命名。提交小说的是“任性的人工智能之我是作家”（简称“我是作家”）团队[2]。

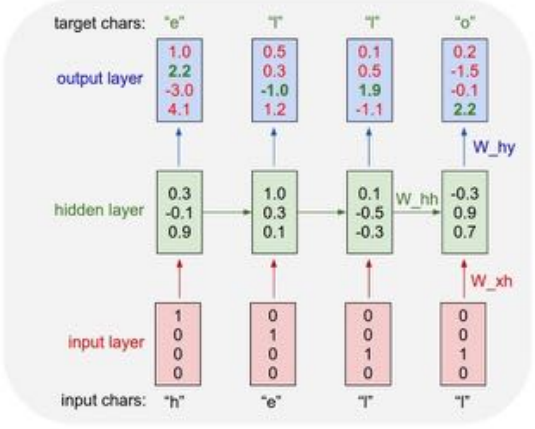
2016年5月，美国多家媒体[3、4]报道，谷歌的人工智能项目在学习了上千本浪漫小说之后写出后现代风格的诗歌。

基于人工智能的文本生成真的已经达到媒体宣传的水平了吗？这些事件背后是怎样的人工智能技术？关于机器人写小说的工作，我们会在另一篇文章《会有那么一天，机器人可以写小说吗？》里进行深入的讨论，他们的工作更多的是基于模板的生成。在这篇文章里，我们主要想通过三篇文章介绍另一大类方法，即基于统计的文本生成。

● 第一篇 令人吃惊的Char-RNN

关于基于深度学习的文本生成，最入门级的读物包括Andrej Karpathy这篇博客[5]。他使用例子生动讲解了Char-RNN（Character based Recurrent Neural Network）如何用于从文本数据集里学习，然后自动生成像模像样的文本。

图一直观展示了Char-RNN的原理。以要让模型学习写出“hello”为例，Char-RNN的输入输出层都是以字符为单位。输入“h”，应该输出“e”；输入“e”，则应该输出后续的“l”。输入层我们可以用只有一个元素为1的向量来编码不同的字符，例如，h被编码为“1000”、“e”被编码为“0100”，而“l”被编码为“0010”。使用RNN的学习目标是，可以让生成的下一个字符尽量与训练样本里的目标输出一致。在图一的例子中，根据前两个字符产生的状态和第三个输入“l”预测出的下一个字符的向量为<0.1, 0.5, 1.9, -1.1>，最大的一维是第三维，对应的字符则为“0010”，正好是“l”。这就是一个正确的预测。但从第一个“h”得到的输出向量是第四维最大，对应的并不是“e”，这样就产生代价。学习的过程就是不断降低这个代价。学习到的模型，对任何输入字符可以很好地不断预测下一个字符，如此一来就能生成句子或段落。



▲ 图一：Char-RNN的原理说明图

Andrej Karpathy还共享了代码[6]，感兴趣的同学不妨下载来试试，效果会让你震惊。Andrej arpathy在底层使用的RNN的具体实现是LSTM（Long-Short Term Memory），想了解LSTM可以阅读[7]，讲得再清楚不过。

研究人员用Char-RNN做了很多有趣的尝试，例如，用莎士比亚的作品来做训练，模型就能生成出类似莎士比亚的句子；利用金庸的小说来做训练，模型就能生成武侠小说式的句子；利用汪峰的歌词做训练，模型也能生成类似歌词的句子来。

在本文一开始提到的[1]，MIT计算机科学与人工智能实验室的博士后Bradley Hayes也正是利用类似的方法开发了一款模仿候任美国总统Donald Trump的推特机器人，叫DeepDrumpf。例如，图二中这个机器人说，“我就是伊斯兰国不需要的。”

▲ 图二：DeepDrumpf推特机器人发布的类似Trump的推文

据作者介绍，他受到一篇模拟莎士比亚的论文启发，以Donald Trump的演讲和辩论（时常大约几个小时）的字幕作为训练语料，使用深度神经网络学习去训练Trump的模型。他也声称，因为有一篇文章调侃Trump的发言只有小学四年级的水平，因而想到用Trump的语料可能是最容易控制的。

这是一个有趣的应用，记者评论称这个机器人也并不是总能写出好的句子，但至少部分是通顺的。其实，风格并不是很难学到，只要使用的训练语料来自同一个人，而这个人的写作或者发言具有辨识度高的特点。

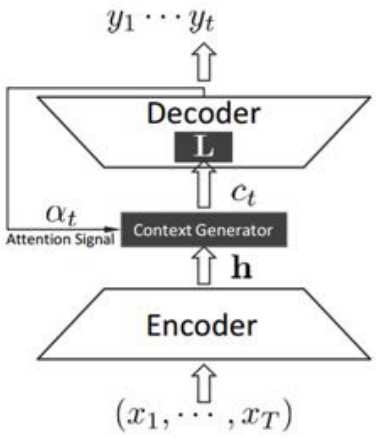
● 第二篇 深度学习生成对话

推荐阅读的第二篇文章是诺亚方舟实验室的尚利峰、吕正东和李航在2015年ACL大会上发表的“Neural Responding Machine for Short-Text Conversation”[9]。大家也许听说过微软小冰，它因为开创性的主要做闲聊（即以娱乐为目的的聊天）式对话，被哈尔滨工业大学的刘挺教授誉为是第二波人机对话的浪潮的代表[8]。小冰的出现也影响到了学术界。除了原来做知识性的问答，一些研究也开始关注闲聊，让机器人和人类搭话，这方面诺亚方舟实验室发表了一系列有影响力的文章。今天介绍的这篇文章在Arxiv.org上发布短短一年时间，已经有67次的引用。

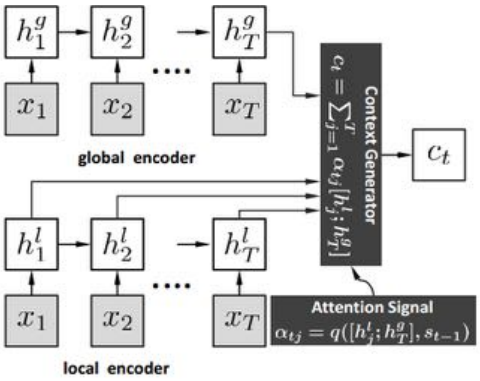
这篇文章[9]尝试用encoder-decoder（编码-解码）的框架解决短文本对话（Short Text Conversation，缩写为STC）的问题。虽然encoder-decoder框架已经被成功应用在机器翻译的任务中，但是对话与翻译不同，对应一个输入文本（post）往往有多种不同的应答（responses）。文中举了一个例子，一个人说“刚刚我吃了一个吞拿鱼三明治”，不同的应答可以是“天哪，才早晨11点”、“看起来很美味哟”或是“在哪里吃的”。这种一对多的情况在对话中很普遍也很自然。的确，不同的人会对同一句话做出不同的反应，即使是同一个人，如果每次回答都一模一样也是很无趣的。

针对这一特点，作者们提出Neural Responding Machine（简称NRM，见图三）框架来解决短文本对话的问题。他们尝试了全局编码和局部编码，最终发现先分别训练，再用图四的结构来做微调训练，效果最佳。全局编码的优点是能够获得全局信息，同样的词在不同情境下会有不同的意义，全局信息可以部分解决这类情况；缺点是，它供给解码的输入比较固定。局部编码利用局部信息，比较灵活多样，刚好可以缓解全局编码的弱点。

这篇论文的另一大贡献是构建了一个比较大的数据集和标注来评价不同的方法。通过对比，所提出的混合全局和局部的方法比以往基于搜索的方法和机器翻译的方法都要好很多。机器翻译的方法生成的句子往往不通顺，得分最低。能比基于搜索的方法好很多也非常不容易，因为基于搜索的方法得到的已经是人使用过的应答，不会不通顺。大家可以在图五的实例中直接感受一下生成的效果。NRM-glo是全局编码的模型，NRM-loc是局部编码的模型，NRM-hyb是混合了全局和局部的模型，Rtr.-based则是基于搜索的方法。



▲ 图三：NRM的基本框架和数据流图



▲ 图四：NRM-hyb方法中的编码器示意图
上下文生成函数是 C_t 是向量 h_j^l 和 h_T^g 的级联

2015到2016年，这篇论文的作者组织了NTCIR-12 STC任务[10]，公开他们的数据集，并提供公共评测。有16个大学或研究机构参加了中文短文本对话任务的评测。2017年，他们将会继续组织NTCIR-13 STC[11]，现已开放注册[12]。除了上一届的基于搜索的子任务，这一次还设立了生成应答的子任务。我们预计今年的结果会更精彩。

ID	Post	NRM-glo	NRM-loc	NRM-hyb	Rtr.-based
1	每年元旦都来高烧。 High fever attacks me every New Year's Day	多休息，增强免疫力！ Rest more to boost your immune system.	每年元旦都去医院看病 Go to see a doctor every New Year's Day	祝早日康复，身体健康！ Get well soon and stay healthy!	高烧… High fever…
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	生日快乐！永远支持你！ Happy birthday! Will stand by you forever!	生日快乐！永远年轻漂亮！ Happy birthday! Wish you always stay young and pretty!	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.	我只是老了一岁 I just gain only one more year.
3	吃素第一天，坚持住，很亲肚。 First day of being a vegetarian. Hold on, **CuiDuoLa**	必须来支持！加油！ I come to offer my support. Keep it up!	坚持住，坚持下去。 Hold on, keep it up.	每天都是素食主义者 I have been being a vegetarian everyday	坚持住！ Hold on!
4	这种天气真应该带个帅锅出去踏青赏花 We should go out with some cute guys to enjoy a great outing in such a nice weather.	这是哪儿啊？真美！ Where is it? It is so beautiful!	阳光明媚，心情好畅。 Such a nice sunny day! I am in a great mood.	这种天气真不是盖的。 It is indeed a very nice weather.	文山啊 出去踏青寻找灵感 哈哈 **WenShan**, let's go out to get some inspiration. Ha! Ha!

▲ 图五：四个实例：针对同一个发文各种方法给出不同的应答

● 第三篇 被媒体误解的谷歌人工智能写诗

第三篇文章是Samuel Bowman等发表在Arxiv.org上的名为“Generating Sentences from a Continuous Space”的文章[13]。作者分别来自斯坦福大学、马萨诸塞大学阿姆斯特分校以及谷歌大脑部门，工作是在谷歌完成的。

这一工作曾被媒体广泛报道，但我发现很多报道(例如[3, 4])都对论文的工作有一些误解。一些记者将图六所示的文字误认为是机器人写出来的后现代风格的诗歌，其实不然。这只是作者在展示他们的方法可以让句子级别的编码解码更连续。具体而言，在他们学习到的空间中，每个点可以对应一个句子，任意选定两个点，例如在图六中，一对点对应的句子分别是“i want to talk to you.”和“she didn’t want to be with him”，两点之间的连线上可以找出间隔均匀的几个点，将它们也解码成句子，会发现，这些句子好像是从第一句逐渐变化成了最后一句。

“ i want to talk to you . ”
“i want to be with you . ”
“i do n’t want to be with you . ”
i do n’t want to be with you .
she did n’t want to be with him .

he was silent for a long moment .
he was silent for a moment .
it was quiet for a moment .
it was dark and cold .
there was a pause .
it was my turn .

▲ 图六：在训练得到的空间中，任意两点之间的路径实例

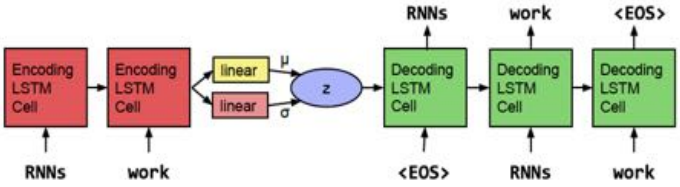
得到这样的结果实属不易。在文章的一开始，作者就给出了一个例子，来说明传统的自动解码并不能很好地编码完整的句子。如图七所示，从句子“i went to the store to buy some groceries”到句子“horses are my favorite animals”，中间取的点经过解码得到的句子呈现在它们之间。可以发现，这些句子未必是符合语法的英文句子。与之相比，图六呈现的句子质量要好很多，不仅语法正确，主题和句法也一致。

i went to the store to buy some groceries .
i store to buy some groceries .
i were to buy any groceries .
horses are to buy any groceries .
horses are to buy any animal .
horses the favorite any animal .
horses the favorite favorite animal .
horses are my favorite animal .

▲ 图七：在训练得到的空间中，任意两点之间的路径实例

这篇文章的想法非常有意思，他们想使用VAE（varation-alautoencoder的简称）学习到一个更连续的句子空间。如图八所示，作者使用了单层的LSTM模型作为encoder（编码器）和decoder（解码器），并使用高斯先验作为regularizer（正规化项），形成一个序列的自动编码器。比起一般的编码解码框架得到的句子编码往往只会记住一些孤立的点，VAE框架学到的可以想象成是一个椭圆形区域，这样可以更好地充满整个空间。我的理解是，VAE框架将贝叶斯理论与深度神经网络相结合，在优化生成下一个词的目标的同时，也优化了跟先验有关的一些目标

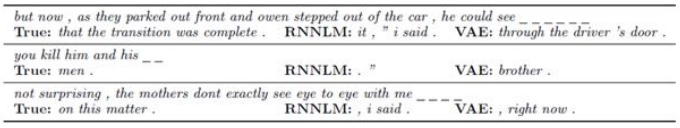
（例如KL cost和crossentropy两项，细节请参考论文），使对一个整句的表达更好。



▲ 图八：VAE（variational autoencoder）语言模型的核心结构

当然，为了实现这一想法，作者做了很多尝试。首先，对图八所展示的结构做一些变形并没有带来明显的区别。但在优化时，使用退火的技巧来降低KL cost和训练时把适当比例的词变为未知词（即word dropout）这两项技术就非常有效。

作者们通过两个有意思的实验来展示了他们的结果。一个是做填空题，如图九所示，隐藏句子的后20%，让模型来生成后面的部分。从几个例子看，VAE的方法比RNN语言模型（简称RNNLM）更加通顺和有信息量。第二个实验就是在两个句子之间做轮移（Homotopy，也就是线性插值），对比图六和图七，可以看出VAE给出的句子更平滑而且正确，这一点可以间接说明学习到的句子空间更好地被充满。



▲ 图九：使用beam search来对句子缺少的部分填空

当然，作者们还给出了一些定量的比较结果。在比较填空结果时，他们使用了adversarial evaluation（对抗评价）。具体的做法是，他们取样50%的完整句子作为正例，再拿50%的由模型填空完成的句子作为负例。然后训练一个分类器，如果一个模型填的越难与正例分开，就说明这种模型的生成效果更好，更具欺骗性。因此，可以认为这一模型在填空任务上更出色。实验的结果也支持VAE比RNNLM更好。

● 问题与难点

人工智能真的会创作吗？使用深度学习技术写出的文章或者对话，的确是会出现训练集合里未见过的句子。例如，一个原句的前半段可能会跟上另一个原句的后半段；也可能除了词，搭配组合都是训练集里没有的。这看起来有些创作的意味，但是细究起来，往往是原句的部分更为通顺和有意义。目前的技术可以拼凑，偶尔出现一两个好玩的点，但是写得长了，读起来会觉得没头没脑，这是因为没有统领全篇的精神，跟人类的作家比当然还是相差很远。

机器学习到的还只是文字表面，没有具备人要写文章的内在动因。人写文章表达的是自己的思想和感受，这是机器所没有的。因此，即使是机器写文章，具体想要表达什么，似乎还要由人来控制。但如果控制得太多，看起来又不那么智能，少了些趣味。我认为，要想让机器更自由地写出合乎逻辑的话来，我们还需要类似VAE那篇文章一样更深入的研究，对句子甚至段落的内在逻辑进行学习。

另外，人在写一篇文章的时候，很容易自我衡量语句是否通顺、思想是否表达清楚以及文章的结构是否清晰有趣，机器却很难做到。因此，优化的目标很难与真正的质量相一致。目前的自然语言理解技术对于判断句法语法是否正确可能还有些办法，但要想判断内容和逻辑上是否顺畅，恐怕还需要常识和推理的帮助，这些部分暂时还比较薄弱。但也并非毫无办法，我相信未来对文本生成的研究一定会涉及这些方面。

期待更多的人来研究如此好玩的文本生成。

● 参考文献

1：http://www.csail.mit.edu/deepdrumpf

2：http://kotoba.nuee.nagoya-u.ac.jp/sc/gw/doc/20160321f.pdf

3：https://www.theguardian.com/technology/2016/may/17/googles-ai-write-poetry-stark-dramatic-vogons

4：http://www.androidauthority.com/google-ai-poetry-692231/

5：http://karpathy.github.io/2015/05/21/rnn-effectiveness/

6：https://github.com/karpathy/char-rnn

7：http://colah.github.io/posts/2015-08-Understanding-LSTMs/

8：刘挺，人机对话浪潮：语音助手、聊天机器人、机器伴侣，中国计算机学会通讯，2015年第10期，2015

9：Lifeng Shang, Zhengdong Lu, Hang Li. Neural Responding Machine for Short Text Conversation. Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP' 15), 1577-1586, 2015.

10：http://ntcir12.noahlab.com.hk/stc.htm

11：http://ntcirstc.noahlab.com.hk/STC2/stc-cn.htm

12：http://research.nii.ac.jp/ntcir/ntcir-13/howto.html

13：Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, Samy Bengio. Generating Sentences from a Continuous Space. arXiv preprint, arXiv:1511.06349, 2015.

培养一个人类医生至少需要八年，那么人工智能呢？

● 一位普通患者的非典型住院之旅

2015年，微软亚洲研究院的资深研究员闫峻在北京一家医院做了一场外科手术，手术前后，他在医院住了十天。但当他躺在病床上，想安安静静看上几本书，或者干脆好好补个长觉时，却不断被各种嘈杂的声音打扰。如果你也曾在医院待过一两天，你几乎不会错过这样的场景：忙不过来的医生和不断追问的家属，手忙脚乱。

闫峻躺在病床上，研究者的好奇心让他试图用技术的眼光寻找原因。闫峻分析：“医生资源十分有限，他没有办法将他的知识表达给所有的患者，而由于信息不对称的存在，患者的安全感是不够的。患者需要得到医生的专业解释和帮助，这样心里才会有底。”

“提高生产力”这个概念几乎对于每一个微软员工都熟捻于心。当这位微软亚洲研究院的研究员结束了他十天的医院之旅，走出医院时，一个利用已有的人工智能技术帮助医院医生提高生产力的计划在他的脑海里也逐渐萌芽。

● 知识挖掘与人工智能

闫峻是微软亚洲研究院数据挖掘与企业智能化组的负责人，今年是他在研究院工作的第10个年头。早在2003年，闫峻在北大读博士的第二年开始就以实习生的身份加入了微软亚洲研究院，彼时亚洲研究院刚成立五年，国内互联网和计算机技术发展迅猛。闫峻在博士期间的研究方向是模式识别，其实就是现在广义的机器学习。

近几年，人工智能俨然成为了计算机领域最火热的研究方向，这背后离不开机器学习、大数据和云计算等的共同作用。社会上关于人工智能的讨论始终不绝于耳，人工智能威胁论、人工智能奇点论等等。但无论媒体如何鼓吹，人工智能领域的研究专家基本上都会给你同样的答案：真正的人工智能离我们还很遥远。

目前，人工智能已经能够在一些如图像识别、语音识别、实时翻译等特定的任务上表现优异，但在更深层次的思考、总结、联想、创造等等方面的进展十分有限。此前计算机战胜人类顶级围棋棋手加强了研究者们用机器学习构建计算机智能的信心，诚然，机器学习目前也被视为是通往人工智能一条近路之一。

事实上，知识的挖掘和知识的表述是人工智能里最基础的问题。我们怎么衡量一个人是“智能”的？一方面，我们会说这个人的脑子很灵活，反应很快。另一方面我们会说这个人有很多经验，懂很多知识。对应到人工智能的研究上来，也就是我们需要人工智能本身拥有很多的背景知识，也需要他有很强的推理能力、理解能力和判断能力。目前，闫峻所在的数据挖掘与企业智能化组主要的研究方向便主要聚焦于知识的挖掘，表示与计算，具体来说就是研究如何从不同的数据来源里挖掘出知识，帮助人工智能的其他算法在具体的场景中完成一些特定的任务。

这一次，闫峻想借用自己的研究成果，为医生培养他们自己的人工智能助手，辅助医生完成一些他们无暇顾及的任务，并为患者提供更多的人文关怀。最终的目的是希望这个人工智能医生助手兼学徒有一天能够顺利出师，能像一个真正的医生一样耐心细致的回答病人的疑惑，协助病人的康复，缓解目前的医疗资源紧缺的局面。

● 培养一个人类医生至少需要八年，那么人工智能呢？

闫峻在病假后的第一次午饭时，就和研究院的同事们聊起了这次住院经历。让他意外的是，几乎每一个同事——甚至比他年纪小很多的90后年轻研究员都表达了强烈的同感，他们的父母或者亲人都或多或少的经历过同样的事情。在那天的午饭上，数据挖掘与企业智能化组的研究员们提出了一个创造性的想法：为医生构建一个人工智能医生助手！

对于人工智能医生助手而言，技术上最关键的一环就是让计算机学会一些必要的医学知识。这恰恰是闫峻博士所在的数据挖掘与企业智能化组所专攻的。闫峻习惯将机器学习知识的过程与人类学习知识的方式相类

比，将这个过程分为四个阶段。“就像指导你的孩子一样学知识。”闫峻形容道。

第一个阶段是人类婴儿出生到上学之前的这个阶段。这个阶段孩子还没有很强的读写和分析问题的能力，更多地是从他的父母或者身边的环境中学习一般性的知识(common sense knowledge)，例如糖是甜的，天是蓝的。这种让计算机学会一般性的知识是较为基础的研究工作，机器会从不同的数据源里面学到一些基本的事实和一些基本的关系。与之对应的，不久前数据挖掘与企业智能化组还发布的Microsoft Concept Graph和Microsoft Concept Tagging模型就试图让计算机掌握部分常识。

第二个阶段则是小孩进入学校开始读书直至大学毕业的阶段。这个过程其实人们主要的学习方式是读书，学习书本上的知识。人们在读书的过程中会不断地分析、总结、归纳，将一些知识进行整合和关联，形成系统性的知识体系，我们称之为知识图谱（knowledge graph）。因此在这个阶段，让机器拥有阅读能力是一个非常重要的课题。研究员们让计算机阅读了大量医学学生的必学书目，让计算机从这些文本中获取作为医生助手最基础的医学知识作为储备，并为下一步进医院“实习”做好准备。

第三阶段则是研究生阶段，你会有一个导师，而你学习的方式已经不再是被动的接受书本知识，而是主动地从他人——你的导师那里学习知识了。这种学习方式中师生间有很多互动，学生会主动询问老师解决办法，从机器学习的角度来理解就是“主动学习”。到了这个阶段，研究员们就将这位人工智能助手送到医院的具体某个科室，选择一位医生作为它的“师傅”。进行简单的“拜师仪式”之后，机器人将成为某位医生的学徒。对于机器学徒不懂得问题，会在聚类总结后向导师主动询问来获取知识。

除了主动提问，跟人类的拜师学艺没有什么区别，计算机向师傅学习的最便捷的方式，就是从医生的匿名诊疗记录入手。这里暗藏着医生的经验知识，在医疗信息化的今天，每一位医生看过的病人信息都会以电子化的形式记录下来，包括病人的症状，各项检查结果以及医生的治疗方案。通过对病例的通读与学习之后，这位虚拟学徒便会对这个科室有一个大体的认识，包括常见症状以及诊疗方案等。但这并不意味着这位人工智能助手可以成功出师了，它还需要最后一个阶段的学习。

最后一个阶段人们通常称之为“增强学习”。很显然，知识的更迭日新月异，人们在走出象牙塔之后发现还需要不断地学习此时的学习更多地会依赖于自己经验的积累以及不断的试错。此前AlphaGo所采用的增强学习就是源自这类思想，它会不断尝试各种下法，观察不同的决策带来的可能性，并以此进行经验总结，从而指导他与真人下棋时的决策。在这个阶段，研究员们才让这位“助手”在监督无风险情况下有选择地与患者进行接触。搜集反馈，提升自己的经验和能力。所有的知识学习阶段都是在“终身学习”的框架下不断迭代的进行。

“培养一个医生至少需要八年，但是计算机几乎不眠不休、不吃不喝，并且过目不忘，所以这个学习过程毫无疑问可以大大缩短。”闫峻博士虽然没有给出一个具体的时长，因为这个很大程度上与人工智能医生所处的科室关系很大，但可以确信的是，培养一个人工智能虚拟医生助手的成本远远低于培养一个真正人类的医生助手。

● 做些医生做不到的事情

那么，人工智能医生助手究竟能帮医生和病患们做些什么呢？

目前，这项关于人工智能医生助手的医疗项目已经与多家医院和医疗机构合作进行试点，由微软亚洲研究院提供算法和解决方案，对接医院需求，医生与研究员一起从事临床实验。研究员们为每个医生各自搭建了属于他们的人工智能助手。这位医生助手会以微信服务号等多种不同的形式协助医生，并贯穿在治疗前、治疗中和治疗后的整个就医体验中。

其实在许多疾病的治疗，特别是外科手术中，大部分患者遇到的问题是很类似的，机器人可以记忆并理解这些常见的问题，辅助医生事先向患者提问，搜集患者的有用信息，并给出初步的诊疗方案供医生参考，甚至能让患者在与医生见面之前提前做好必备的检查，这样就能大大缩短病人

实际面见医生的时间和流程，在有限的时间里，直接切入关键问题。

除了治疗前的提前沟通和治疗途中的有问必答之外，人工智能助手还能做些医生很难细致的面面俱到的事，例如患者随访，收集医疗大数据，增进对病患的人文关怀。

“在和医生合作的过程中，我们发现医生看病的过程跟机器学习也十分相似。”闫峻博士说。患者介绍病情即提供了机器学习算法的特征维度，接下来，医生根据患者的病情提供治疗方案，即进行预测，这很像训练机器学习的一个算法模型。但最重要的是，医生还需要了解这个治疗方案是否合适、患者有无并发症等问题，如果没有这一反馈，就无法得知医生的预测结果是否成功、没有机器学习需要的重要的数据标注，也就很难用此数据提高整体医疗水平。

“这其实就是一个患者随访的过程。”闫峻说道。

闫峻在这次与医院的合作过程中认识了不少医生朋友。他了解到，每位医生每年的患者量非常巨大，对每一位患者都进行长期跟踪随访几乎是不可能的，“有的医生平均一年可能做上千台手术”。从学术科研的角度来看，真实有效随访的医疗大数据量很少，影响医疗水平的提高；而从患者的角度来说，由于离开医院后很难得到后续的提醒和关怀，也影响整体的就医感受。

微软亚洲研究院的研究员们也很快为这个问题提供了解决方案。研究员为医生们在各自的人工智能医生助手提供了一个后台，医生能在机器人的后台上设置一些规则、随访问题，例如治疗效果如何、是否有并发症等。在患者给出回复后，计算机便会通过自然语言处理技术来理解患者的回答，并将其变成知识库中的一部分，用来做数据分析。当患者有不理想的治疗结果反馈时，机器人也会及时地将这一信息通知医生。这样医生就能第一时间联系病人进行进一步的治疗，杜绝潜在的风险。特别是对于外地来省会医院等治疗的病人，由于路途遥远，这种自动化定期随访将方便他们与医生构建长期联系的机制。

“这个医疗项目本质上是提升医生和患者之间的沟通效率，让两者能够相互信赖。”闫峻博士总结道。通过人工智能医生助手，患者在就诊前就可以事先和机器人做一些简单的沟通，大概了解可能的治疗方案，什么时间去医院比较合适，在本地治疗还是需要去更高级的医院。患者能够通过这种前期沟通的方式一定程度上消除这种信息的不对等，患者在治疗过程中有任何疑惑也能得到及时的回答。在治疗之后，患者还能得到医院的长期关怀，研究员们还为这个机器人加入了一些很人性化的功能，例如提醒病人什么时候需要吃药，吃药的剂量；或者在患者生日和特殊节日时，发送祝福，祝愿患者身体健康等。这些对于计算机而言十分简单的任务，对于工作繁重的医生来说却几乎不可能面面俱到的完成。而从医院的角度来看，也可以收集真实的医疗大数据，帮助医院提升整体的医疗水平和科研水平。

● 最大的困难是没有困难

困难重重。这个词可能是很多人第一次接触到这个项目时的想法。但出人意料的是，闫峻博士却不觉得这个项目有什么“最大的困难”。

对于学术界的人来说，如果想做这样一个项目，最大的困难可能来源于对资金的支持。而对于工业界的人来说，你很难说服自己的老板去做一个和整个公司的商业目标不是完全契合的项目。但这两种困难，对于微软亚洲研究院这样一个十几年如一日深耕计算机基础科研的机构来说，是几乎不存在的。

“我觉得在这里做研究最好的事情是我们没有太多的后顾之忧，也没有什么资金的压力。你只需要证明你做的这件事情本身是有价值的，那就没有人会阻拦你。如果说困难，困难都是研究上面的具体问题，例如数据量太大导致训练算法的过程耗时太久等等。但这种问题是每一个做研究的人都会碰到的。”

类似于人工智能医生助手这类的研究项目在微软还有很多。帮助视觉障碍患者重新“看见”周围世界的Seeing AI项目，帮助渐冻症患者用眼睛驾驶轮椅的Eye Gaze Wheelchair项目，用机器学习技术寻找艾滋病毒的核心蛋白计划，利用计算机视觉和机器学习识别癌症病理切片……正是因为有这么多心怀善意与正能量的酷炫极客们正在用技术来改善身边每一个人的计算体验，这个世界才会越来越美好。

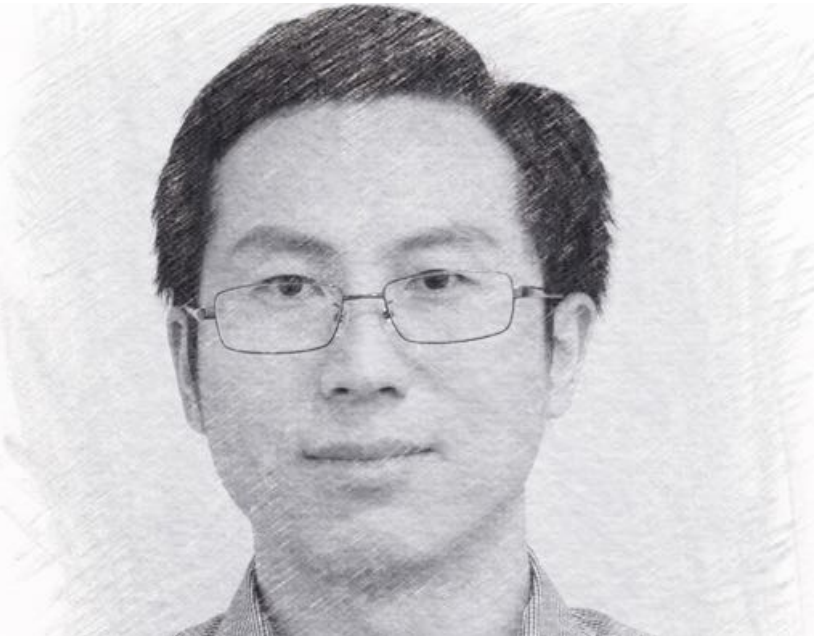
目前，这个人工智能医疗助手项目还处于试验的早期阶段，相关的技术还在根据试验的结果不断完善。至于下一步的计划，闫峻博士表示：“希望能将这项技术真正应用到临床上，让每一个医疗环节都能多一份信赖。”

与此相关，为了更好的了解医生和患者的实际需求，也为了有更多的人工智能技术能够服务于中国的医疗健康问题，微软亚洲研究院还与中文信息学会语言与知识计算专业委员会联合组织了知识图谱与智慧医疗论坛。期待专家学者们能够为医疗信息化、精准医疗等问题提供更高效，更贴近民生的解决方案！

阅读“文本型医疗大数据，拿来就可用？”
扫描二维码即可查看

对偶学习：一种新的机器学习范式

作者： 微软亚洲研究院主管研究员 秦涛



● 问题：如何降低对大规模标注数据的依赖性

近年来，以深度学习为代表的人工智能技术取得了长足的进步，在很多应用取得了巨大的成功。在图像识别中，深度学习已经超过了人类的水平，最近微软刚刚宣布在语音识别中也超过了人类的平均水平，在机器翻译中基于深度学习的模型也超过了以前的统计机器翻译，另外基于深度学习的阿法狗也打败了围棋世界冠军李世石。

深度学习之所以能够取得巨大的成功，一个非常重要的因素就是大数据，特别是大规模的带标签的数据。例如在图像识别中，深度神经网络使用上百万的带标签的图像进行训练，在机器翻译中我们会用上千万的双语句对进行训练，在围棋中我们会用上千万的专业棋手的落子进行训练……这种做法有两个局限性。首先，人工标注获取标签的代价很高。例如我们考虑机器翻译这个任务：现在市场人工翻译一个单词的价格差不多是5到10美分，如果一个句子的平均长度为三十个单词，那么1000万个句子人工翻译的代价差不多是7.5美分×30×1000万，约等于2200万美元。现在一个商业公司的翻译引擎通常支持上百种语言的相互翻译，为了训练这样规模的翻译模型，人工标注的代价就会达到上千亿美元。其次，在很多任务中，我们没办法收集到大规模的标注数据，例如在医疗中或在小语种的相互翻译。为了使深度学习能够取得更广泛的成功，我们需要降低其对大规模标注数据的依赖性。为了解决这个问题，我们提出了一种新的学习范式，我们把它称作对偶学习。

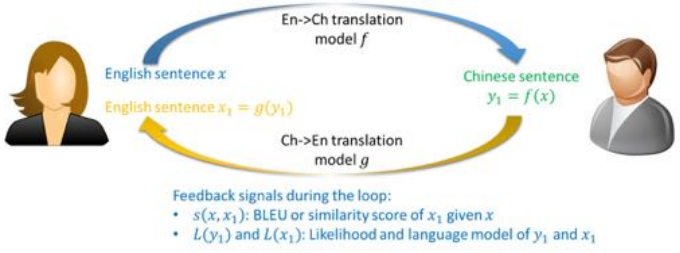
● 另辟蹊径，将未标注数据用起来

我们注意到，很多人工智能的应用涉及两个互为对偶的任务，例如机器翻译中从中文到英文的翻译和从英文到中文的翻译互为对偶、语音处理中语音识别和语音合成互为对偶、图像理解中基于图像生成文本和基于文本生成图像互为对偶、问答系统中回答问题和生成问题互为对偶，以及在搜索引擎中给检索词查找相关的网页和给网页生成关键词互为对偶。这些互为对偶的人工智能任务可以形成一个闭环，使从没有标注的数据中进行学习成为可能。

对偶学习的最关键一点在于，给定一个原始任务模型，其对偶任务的模型可以给其提供反馈；同样的，给定一个对偶任务的模型，其原始任务的模型也可以给该对偶任务的模型提供反馈；从而这两个互为对偶的任务可以相互提供反馈，相互学习、相互提高。

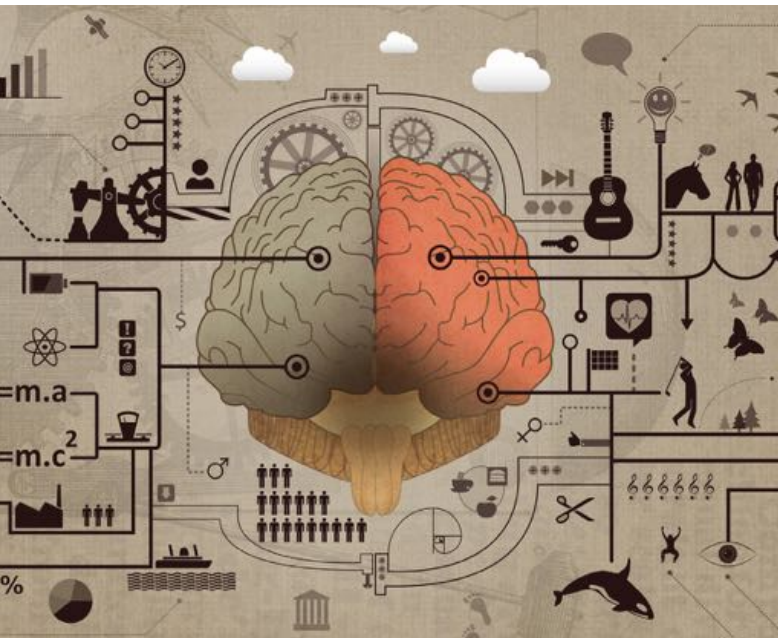
下面我们还是以机器翻译为例子来说明对偶学习的基本思想。

考虑一个对偶翻译游戏，里面有两个玩家小明和爱丽丝，如下图所示。小明只能讲中文，爱丽丝只会讲英文，他们两个人一起希望能够提高英文到中文的翻译模型f和中文到英文的翻译模型g。给定一个英文的句子x，爱丽丝首先通过f把这个句子翻译成中文句子y1，然后把这个中文的句子发给小明。因为没有标注，所以小明不知道正确的翻译是什么，但是小明可以知道，这个中文的句子是不是语法正确、符不符合中文的语言模型，这些信息都能帮助小明大概判断翻译模型f是不是做的好。然后小明再把这个中文的句子y1通过翻译模型g翻译成一个新的英文句子x1，并发给爱丽丝。通过比较x和x1是不是相似，爱丽丝就能够知道翻译模型f和g是不是做得好，尽管x只是一个没有标注的句子。因此，通过这样一个对偶游戏的过程，我们能够从没有标注的数据上获得反馈，从而知道如何提高机器学习模型。



▲ 对偶翻译游戏

实际上这个对偶游戏和强化学习的过程比较类似。在强化学习中，我们希望提高我们的策略以最大化长远的回报，但是没有标注的样本告诉我们在某个状态x哪个动作y是正确的。我们只有通过使用这个策略在不同的状态下执行不同的动作，观测该动作带来的回报，从而改善我们的策略。

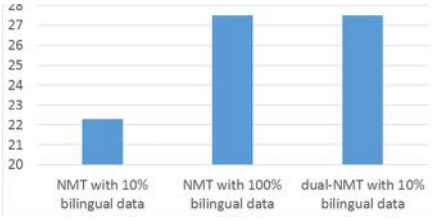


在以上这个翻译对偶游戏中，两个翻译模型就是我们的策略，因为没有标注的双语句对，所以我们不能直接改善它们。这个对偶游戏把一个没有标注的句子x，先翻译成另外一种语言的句子y1，再翻译回来为x1，这里x就是强化学习中的状态，y1和x1就是我们的策略所执行的动作，x和x1的相似度就是我们获得的回报。

我们可以用已有的强化学习的算法来训练我们这两个翻译模型，比如策略梯度方法。策略梯度方法的基本思想非常简单：如果我们在执行某个动作之后，观测到了一个很大的回报，我们就通过调整策略（在当前策略函数的参数上加上它的梯度）来增加这个状态下执行这个动作的概率；相反，如果我们在执行某个动作之后，观测到了一个很小的回报，甚至是负的回报，那么我们就需要调整策略（在当前策略函数的参数上减去它的梯度），以降低在这个状态下执行这个动作的概率。

● 实测与展望

我们在一个英语-法语翻译的公共数据集上测试了对偶学习的有效性，并和当前最好的算法（NMT，基于深度神经网络算法）进行了对比。这个数据集的训练集有1200万个英法的双语句对，NMT用了所有的1200万个双语句对训练模型。我们的对偶算法只用了不到10%的双语句对初始化两个翻译模型f和g，再用没有标注的英法语句进行训练。如下图所示，我们的对偶学习算法只用10%的标注数据就达到了和NMT用上所有标注数据而取得的相似的甚至更好的翻译准确度，也就是说我们可以把数据标注的代价从2000万美元降到200万美元左右。这个结果说明了我们提出的对偶学习方法能够非常有效地利用未标注的数据。



▲ 对偶学习算法和NMT算法的有效性对比

对偶学习的基本思想是两个对偶的任务能形成一个闭环反馈系统，使我们得以从未标注的数据上获得反馈信息，进而利用该反馈信息提高对偶任务中的两个机器学习模型。该思想具有普适性，可以扩展到多个相关任务上面，前提是只要它们能形成一个闭环反馈系统。例如，从中文翻译到英文，然后从英文翻译到日文，再从日文翻译到中文。另外一个例子是从图片转化成文字，然后从文字转成语音，再从语音转化成图片。

对偶学习（dual learning）和已有的学习范式有很大的不同。首先，监督学习（supervised learning）只能从标注的数据进行学习，只涉及一个学习任务；而对偶学习涉及至少两个学习任务，可以从未标注的数据进行学习。其次，半监督学习（semi-supervised learning）尽管可以对未标注的样本生成伪标签，但无法知道这些伪标签的好坏，而对偶学习通过对偶游戏生成的反馈（例如对偶翻译中x和x1的相似性）能知道中间过程产生的伪标签（y1）的好坏，因而可以更有效地利用未标注的数据。我们甚至可以说，对偶学习在某种程度上是把未标注的数据当作带标签的数据来使用。第三，对偶学习和多任务学习（multi-task learning）也不相同。尽管多任务学习也是同时学习多个任务共的模型，但这些任务必须共享相同的输入空间，而对偶学习对输入空间没有要求，只要这些任务能形成一个闭环系统即可。第四，对偶学习和迁移学习（transfer learning）也很不一样。迁移学习用一个或多个相关的任务来辅助主要任务的学习，而在对偶学习中，多个任务是相互帮助、相互提高，并没有主次之分。因此，对偶学习是一个全新的学习范式，我们预计其会对机器学习领域产生很大的影响，特别是考虑到以下两个点。

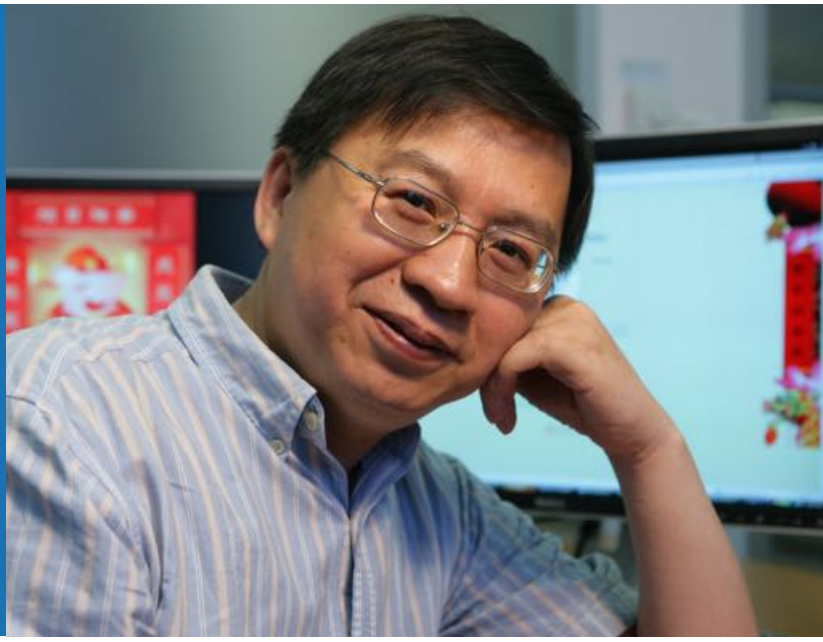
第一，很多深度学习的研究人员认为人工智能和深度学习的下一个突破是从未标注的数据进行学习。由于未标注的数据没有标签信息，无监督学习主要聚焦在聚类、降维等问题，因此，如何从未标注的数据进行端到端的学习（把数据映射到标签，如分类、预测等任务）一直没有很好方法。我们提出的对偶学习提供了一个利用未标注数据进行端到端学习的有效方式。

第二，近年来强化学习取得了很大的成功，吸引了众多的研究人员和工业界人员的关注。但是到目前为止，强化学习的成功主要是在各种游戏上，因为在游戏中规则定义得非常清楚，并且很容易通过玩大量的游戏获取回报信息（例如Atari游戏中每一步的得分，围棋中最后的输赢），从而改善游戏策略。然而在游戏以外的实际应用中并没有规范定义的规则，也很难获得回报信息（或者获取的代价很高），因而强化学习在复杂的实际应用中的成功还是比较有限。而对偶学习提供了一种为强化学习获取奖励信息的方式，并证实了强化学习在复杂应用（如翻译）中成功的可能。

相关论文
扫描二维码即可查看

如何在NLP领域干成第一件事

作者： 微软亚洲研究院副院长 周明



自然语言处理（简称NLP），是研究计算机处理人类语言的一门技术，包括：

1.句法语义分析：对于给定的句子，进行分词、词性标记、命名实体识别和链接、句法分析、语义角色识别和多义词消歧。

2.信息抽取：从给定文本中抽取重要的信息，比如，时间、地点、人物、事件、原因、结果、数字、日期、货币、专有名词等等。通俗说来，就是要了解谁在什么时候、什么原因、对谁、做了什么事、有什么结果。涉及到实体识别、时间抽取、因果关系抽取等关键技术。

3.文本挖掘（或者文本数据挖掘）：包括文本聚类、分类、信息抽取、摘要、情感分析以及对挖掘的信息和知识的可视化、交互式的表达界面。目前主流的技术都是基于统计机器学习的。

4.机器翻译：把输入的源语言文本通过自动翻译获得另外一种语言的文本。根据输入媒介不同，可以细分为文本翻译、语音翻译、手语翻译、图形翻译等。机器翻译从最早的基于规则的方法到二十年前的基于统计的方法，再到今天的基于神经网络（编码-解码）的方法，逐渐形成了一套比较严谨的方法体系。

5.信息检索：对大规模的文档进行索引。可简单对文档中的词汇，赋之以不同的权重来建立索引，也可利用1、2、3的技术来建立更加深层的索引。在查询的时候，对输入的查询表达式比如一个检索词或者一个句子进行分析，然后在索引里面查找匹配的候选文档，再根据一个排序机制把候选文档排序，最后输出排序得分最高的文档。

6.问答系统： 对一个自然语言表达的问题，由问答系统给出一个精准的答案。需要对自然语言查询语句进行某种程度的语义分析，包括实体链接、关系识别，形成逻辑表达式，然后到知识库中查找可能的候选答案并通过一个排序机制找出最佳的答案。

7.对话系统：系统通过一系列的对话，跟用户进行聊天、回答、完成某一项任务。涉及到用户意图理解、通用聊天引擎、问答引擎、对话管理等技术。此外，为了体现上下文相关，要具备多轮对话能力。同时，为了体现个性化，要开发用户画像以及基于用户画像的个性化回复。

随着深度学习在图像识别、语音识别领域大放异彩，人们对深度学习在NLP的价值也寄予厚望。再加上AlphaGo的成功，人工智能的研究和应用变得炙手可热。自然语言处理作为人工智能领域的认知智能，成为目前大家关注的焦点。很多研究生都在进入自然语言领域，寄望未来在人工智能方向大展身手。但大家常常遇到一些问题——俗话说“万事开头难”，如果第一件事情成功了，学生就能建立信心，找到窍门，今后越做越好。否则，也可能就灰心丧气，甚至离开这个领域。这里针对给出我个人的建议，希望我的这些粗浅观点能够引起大家更深层次的讨论。

● 建议1：如何在NLP领域快速学会第一个技能？

我的建议是：找到一个开源项目，比如机器翻译或者深度学习的项目。理解开源项目的任务，编译通过该项目发布的示范程序，得到与项目示范程序一致的结果。然后再深入理解开源项目示范程序的算法。自己编程实现一下这个示范程序的算法。再按照项目提供的标准测试集测试自己实现的程序。如果输出的结果与项目中出现的结果不一致，就要仔细查验自己的程序，反复修改，直到结果与示范程序基本一致。如果还是不行，就大胆给项目的作者写信请教。在此基础上，再看看自己能否进一步完善算法或者实现，取得比示范程序更好的结果。

● 建议2：如何选择第一个好题目？

工程型研究生，选题很多都是老师给定的。需要采取比较实用的方法，扎扎实实地动手实现。可能不需要多少理论创新，但是需要较强的实现能力和综合创新能力。而学术型研究生需要取得一流的研究成果，因此选题需要有一定的创新。我这里给出如下的几点建议。

先找到自己喜欢的研究领域。你找到一本最近的ACL会议论文集（编者按：可以参考ACL2016大会综述），从中找到一个你比较喜欢的领域。在选题的时候，多注意选择蓝海的领域。这是因为蓝海的领域，相对比较新，容易出成果。

充分调研这个领域目前的发展状况。包括如下几个方面的调研：方法方面，是否有一套比较清晰的数学体系和机器学习体系；数据方面，有没有一个大家公认的标准训练集和测试集；研究团队，是否有著名团队和人士参加。如果以上几个方面的调研结论不是太清晰，作为初学者可能不要轻易进入。

在确认进入一个领域之后，按照建议一所述，需要找到本领域的开源项目或者工具，仔细研究一遍现有的主要流派和方法，先入门。

反复阅读本领域最新发表的文章，多阅读本领域牛人发表的文章。在深入了解已有工作的基础上，探讨还有没有一些地方可以推翻、改进、综合、迁移。注意做实验的时候，不要贪多，每次实验只需要验证一个想法。每次实验之后，必须要进行分析存在的错误，找出原因。

对成功的实验，进一步探讨如何改进算法。注意实验数据必须是业界公认的数据。

与已有的算法进行比较，体会能够得出比较一般性的结论。如果有，则去写一篇文章，否则，应该换一个新的选题。

● 建议3：如何写出第一篇论文？

接上一个问题，如果想法不错，且被实验所证明，就可开始写第一篇论文了。

确定论文的题目。在定题目的时候，一般不要“……系统”、“……研究与实践”，要避免太长的题目，因为不好体现要点。题目要具体，有深度，突出算法。

写论文摘要。要突出本文针对什么重要问题，提出了什么方法，跟已有工作相比，具有什么优势。实验结果表明，达到了什么水准，解决了什么问题。

写引言。首先讲出本项工作的背景，这个问题的定义，它具有什么重要性。然后介绍对这个问题，现有的方法是什么，有什么优点。但是（注意但是）现有的方法仍然有很多缺陷或者挑战。比如（注意比如），有什么问题。本文针对这个问题，受什么方法（谁的工作）之启发，提出了什么新的方法并做了如下几个方面的研究。然后对每个方面分门别类加以叙述，最后说明实验的结论。再说本文有几条贡献，一般写三条足矣。然后说说文章的章节组织，以及本文的重点。有的时候东西太多，篇幅有限，只能介绍最重要的部分，不需要面面俱到。

相关工作。对相关工作做一个梳理，按照流派划分，对主要的最多三个流派做一个简单介绍。介绍其原理，然后说明其局限性。

然后可设立两个章节介绍自己的工作。第一个章节是算法描述。包括问题定义，数学符号，算法描述。文章的主要公式基本都在这里。有时候要给出简明的推导过程。如果借鉴了别人的理论和算法，要给出清晰的引文信息。在此基础上，由于一般是基于机器学习或者深度学习的方法，要介绍你的模型训练方法和解码方法。第二章就是实验环节。一般要给出实验的目的，要检验什么，实验的方法，数据从哪里来，多大规模。最好数据是用公开评测数据，便于别人重复你的工作。然后对每个实验给出所需的技术参数，并报告实验结果。同时为了与已有工作比较，需要引用已有工作的结果，必要的时候需要重现重要的工作并报告结果。用实验数据说话，说明你比人家的方法要好。要对实验结果好好分析你的工作与别人的工作的不同及各自利弊，并说明其原因。对于目前尚不太好的地方，要分析问题之所在，并将其列为未来的工作。

结论。对本文的贡献再一次总结。既要从理论、方法上加以总结和提炼，也要说明在实验上的贡献和结论。所做的结论，要让读者感到信服，同时指出未来的研究方向。

参考文献。给出所有重要相关工作的论文。记住，漏掉了一篇重要的参考文献（或者牛人的工作），基本上就没有被录取的希望了。

写完第一稿，然后就是再改三遍。

把文章交给同一个项目组的人士，请他们从算法新颖度、创新性和实验规模和结论方面，以挑剔的眼光，审核你的文章。自己针对薄弱环节，进一步改进，重点加强算法深度和工作创新性。

然后请不同项目组的人士审阅。如果他们看不明白，说明文章的可读性不够。你需要修改篇章结构、进行文字润色，增加文章可读性。

如投ACL等国际会议，最好再请英文专业或者母语人士提炼文字。

机器会产生自我意识吗？

作者：微软全球资深副总裁 微软亚太研发集团主席兼微软亚洲研究院院长 洪小文

编者按：

计算机领域的热点总是在不断更替，从大数据到云计算再到人工智能，这些热点的背后离不开专家学者们在这些领域一点一滴聚沙成塔的技术突破。关于人工智能，我们见证了近年来它从默默无闻到炙手可热的过程。继去年《我们需要什么样的机器人》之后，微软全球资深副总裁、微软亚太研发集团主席兼微软亚洲研究院院长洪小文再一次亲手撰写文章，与我们进一步分享了他对人工智能的见解与洞察，归纳起来就是《HI+AI：人机协同 赋能未来》系列。在这一系列三篇文章里，洪小文博士将与读者交流AlphaGo战胜李世石这一标志性事件背后的意义，深入浅出地与我们探讨人类智能与人工智能的区分与联系，以及我们应当如何看待人类与人工智能的关系。

《机器会产生自我意识吗？》是《HI+AI：人机协同 赋能未来》三篇系列文章的下篇。在本文中，洪小文博士以“中文房间”为例子告诉我们，现阶段的人工智能还很弱，离实现强人工智能还很远。而在眼下，HI+AI、人类创意无限的右脑+机器的最强左脑才是赋能未来的最佳路径，最终实现人类与机器的共进化Co-Evolution。阅读其它两面文章请扫描文末二维码。

大脑（Brain）、心灵（Mind）、意识（Consciousness），这是一回事吗？

我认为不是。

在我看来，大脑是一种物理存在，心灵和意识则看不见、摸不着，只能自我感知。

“我”这个词来自于心灵对自身的一种定位和认同，心灵决定了“我”是怎样一个人、为什么和别人会有不一样的反应，意识则像时刻都在流动的运算程序，具有工具性质，它们都寄存在大脑和神经等物理存在里。

关键是，即使今天我们已经有本事上天下海、还能教机器学会很多事情，我们对人类自身的心灵和意识的了解却依然有限，未解之谜还有很多。比如，当医生把一个人的大脑和身体分开，这个人的心灵和意识会有什么变化吗？如果把这个人的大脑移植给另一个人呢？——在后者的“新”大脑里储存的那个“我”，究竟是谁？

又比如，当我们处在睡梦中，意识却可能继续剧烈活动，梦中的“我”甚至登上了火星，这时候，我们的意识是留存在大脑里，还是真的光速般去了火星？

纽约州立大学心理学家小戈登·盖洛普（Gordon Gallup Jr.）主持的“镜子测试”表明，猩猩、大象和海豚似乎能认出镜中映像是自己，这说明这些动物已聪明到具有自我意识，那它们也有心灵吗，也有坚定不移的“我”的定义吗？有没有虔诚的信念和顽固的怀疑？相比之下更加常见的猫和狗平素也有相当不俗的智力表现，却常常搞不清镜中映像究竟是不是外来动物而把自己吓一跳，如果不能确定它们是否也有明确的自我意识，它们又是怎样界定自身和主人的相对关系呢？

最近我在工作之余拜读了耶鲁大学计算机科学教授大卫·格兰特（David Gelernter）的新书《心灵潮汐：揭示意识光谱》（The Tides of Mind: Uncovering the Spectrum of Consciousness），感悟颇深。在书中，格兰特教授援引了马塞尔·普鲁斯特、弗兰兹·卡夫卡、弗拉基米尔·纳博科夫、欧内斯特·海明威等许多作家的故事为例证，解析了创造力的潮汐状周期性特征：当人们处在光谱的上端，也就是意识的高潮期，就会聚焦于外部世界、侧重逻辑推理和经验记忆，而当人们落在光谱的下端，即意识的低潮期，就会偏向于内视心灵、在交叉混乱的叙事脉络与梦境漫游般的狂欢中捕捉创意的火花。

创造力确实并不仅仅在专注和冷静的状态下爆发。杜甫在《饮中八仙

歌》里说：“……李白斗酒诗百篇，长安市上酒家眠……张旭三杯草圣传，脱帽露顶王公前，挥毫落纸如云烟。”若照常理判断，喝酒会影响诗人与书法家的创作状态，但在现实中，诗仙和草圣却因为饮酒而进入到光谱下端，实现了创意的超常发挥。

这样的案例还有很多：在完全失聪的情况下，贝多芬创作出不朽的经典《第九交响曲》；深陷精神失常的痛苦深渊，梵高创作出《向日葵》、《星月夜》等传世名作。1864年冬，潜心研究苯结构而没有进展的德国化学家凯库勒在对着炉火打盹时，梦见咬着自己尾巴的蛇在他眼前转动。醒来后，他写出了苯的结构式，恰恰就是首尾相连一个环……可见，在受到某些特定因素影响时，大脑——特别是右脑——活动很不寻常，这当然不一定都有助于创意的发挥（更多时候应该只是胡思乱想），却也有可能迸发出一些在正常情况下藏得很深、从未闪现的灵感。

我们当然可以好像很科学地说，当某些看似完全无关、并不搭界的思维被人的脑桥意外接在一起，创造力也由此而来。但这“科学”程度也只是好像而已，人类的精神世界到目前为止仍是一片远在我们掌握之外而因此继续值得我们保持敬畏的领域。

无所不在的创造力使人们无论在意识的高潮期或是意识的低潮期常常都能有所发现而收之桑榆，人类教师在传道授业解惑之际也会注意因才施教，遇到某位学生不太能理解自己讲授的知识就会考虑换一种教法，设法让学生达至豁然开朗之境界。若是换成机器来教课，恐怕只能依循预设的程序来“灌输”信息，至于随机应变、触类旁通、举一反三的能力，那是没有的。

继续推论——首先，今天的人工智能仍旧很弱。“中文房间”（The Chinese room argument）是上世纪八十年代由美国哲学家约翰·塞尔（John Searle）提出的一项思维实验。假设在一个密闭的房间里有一个人，只会中文的你仅能通过小纸条和房间里的人交流。当你发现传回来的纸条上，你的每个问题都得到了语意确切的中文回复，这时你还会怀疑房间里的人是否懂中文吗？事实却是，房间里的人根本不懂中文，他拥有的不过是一套假设涵盖了所有中文问答的工具，包括中文字条（数据库）和如何使用这些字条的手册（程序），通过这套工具就能正确找到你要的回答而“回应”你的问题，却同时对谈话内容一无所知。

“中文房间”实验很好地说明了计算机的工具性本质：计算机虽能完成一些特定的任务，但并不代表计算机就能理解这些任务——今天的AI几乎都属于这一类。

更直白地说，无论是Skype Translator实时语音翻译，还是Watson，归根结底都属于弱人工智能（Weak AI）范畴，不是基于对信号和数据意义的理解而生——它们就像“中文房间”中拥有工具的人，只是信息的处理者，却无法真正理解接收到的信息，更谈不上拥有发展出意识的潜能。

对弱人工智能有两种定义，一是聚焦单一任务本身——所以它又称为狭义的人工智能（Narrow AI），二是只有结果没有理解。不过，弱人工智能其实并不弱，它也是很有价值、有益于用户的。若能与传感器网络、大数据等、云计算等技术结合，弱人工智能还是可以具备超越人类的某些能力，因为它本身就是一个专家系统，实际上在经济、科技、民生等各领域都大有可为。

其二，目前看来，强人工智能仍旧遥不可及。我在卡内基梅隆大学的导师罗杰·瑞迪（Dabbala Rajagopal “Raj” Reddy）曾任卡内基梅隆机器人研究所创始主任和美国人工智能协会主席，主攻人工智能和机器交互研究，开发出全球第一台具有连续语音识别能力的系统，因在大规模人工智能系统之设计与构建的先驱性贡献而于1994年获得图灵奖。

瑞迪教授的老师是约翰·麦卡锡（John McCarthy）——1956年，在美国新罕布夏州的达特茅斯学院举办的一场学术会议上，约翰·麦卡锡首次提出了人工智能（Artificial Intelligence）的概念，因而被后辈学者尊称为“人工智能之父”。

我的两位老师都认为，今天业界所做的“人工智能”研究与当年他们的创新思路是有分歧的。若按当前的主流路线，弱人工智能只能在专用的、受限制的轨道上越走越远，若要人工智能由弱变强，还得回到认知论的经典道路上。

与弱人工智能相比，强人工智能（Strong AI）是真的能够理解信号与数据的意义，并由此具备人类所有或大多数的能力。但是这种通用的智能其实也可能很弱，就像人类很难同时对所有的事情都精通，也不可能一直保持着意识的高潮，通用人工智能可能具备很多能力，但每一样都很弱。但另一方面，正是因为这些不完美而让人类的存在如此微妙，并在不完美中迸发出各种奇思妙想。

有人问过约翰·麦卡锡，强人工智能何时才能成真，他的回答很耐人寻味：也许5年，也许500年——我的看法是，在理论基础尚不完备、对人的意识和创造力尚未充分了解的前提下，在架构与平台暂无标准、应用探索刚刚开始的前提下，恐怕真需要等上500年，强人工智能的萌芽才会钻出土壤。

毕竟，人的大脑、心灵、意识是如此的玄妙，以至于人类自己或许永远也搞不清楚，更何况是无机体的机器呢？

对此，大卫·格兰特的观点是，计算机可以模拟理性思考，但很难具备真正的意识，意识不属于计算机，只属于有机体。我认同这个观点。心灵、意识和创造力有极大的关系，很多人的创新来自激情而不是理性。这就意味着我们几乎不可能造出拟人的、有意识的机器。从算法、意识到创意，人工智能若要向更高的智力与智慧阶梯跃进，就要踏踏实实一步步跨越这些障碍。

其三，人工智能并不可怕。计算机是有史以来人类最伟大的发明创造之一，它可以被赋予人类发明的不同算法，进而不断掌握新的能力——很强大不是吗？但更强大的人类创造物也还有一些，比如核武器和航天飞机。所有这些机器都可以被人类利用做各种不同的事，至于究竟是好事还是坏事，只由其使用者的心灵和意识决定——科幻小说作家阿西莫夫提出的“机器人三定律”，从表面上看像是用来约束未来的人工智能机器人，但究其实质，每一条何尝不是对人类创造者的警示？

我们应该担心的，不是人工智能将会强大到颠覆人类统治的地步，而是即便人工智能被越来越多业界领袖企业寄予厚望、不吝资源而加速研发，但迄今为止相关领域的技术进步仍不足以孵化出真正拥有创造力、能够通过自我学习解决未知问题、提出新思维的机器。不过，令人欣慰的是，人工智能正在与机器学习和大数据构成一个足以改变未来的技术“铁三角”；随着我们的探索趋向纵深，在不久的将来，这个“铁三角”将有可能像PC+互联网那样彻底改变人类的生活格局和商业版图。

总而言之，人类创意无限的右脑+机器的最强左脑，这才是赋能未来的最佳路径。唯有潜力无穷尽的人脑才能催生出更先进的算法和能力进化更快的“聪明机器”，人类也因为有了聪明机器的帮助而不断进化，加深着对自身的了解、对计算机技术的理解，以及对更多未知领域的探索，从而实现人类与机器的共进化（Co-Evolution）。

HI+AI：人机协同 赋能未来（上篇）：
机器棋手背后的人脑实力

HI+AI：人机协同 赋能未来（中篇）：
计算机是最好的左脑

17位微软科学家带你预测2017和2027

计算领域为什么缺乏女性？人们经常提到的一个原因就是：缺乏能够激励女孩在科学（Science）、技术（Technology）、工程（Engineering）和数学（Mathematics）领域（简称STEM）追寻梦想的职业榜样。为此，我们特别邀请来自微软全球研究院的17位女性对2017年各自研究领域的前景进行预测，试图打消对科技行业女性的这一偏见。在本文中，我们请她们分享了她们对未来10年的展望。

张冬梅 微软亚洲研究院副院长

● 2017年，数据分析与可视化领域将有哪些重大进展？

2017年，数据分析与可视化领域的重大技术突破将出现在智能数据发现（Smart Data Discovery)方向，其核心就是自动、快速地从数据中挖掘出用户感兴趣的、最有价值的信息，并利用交互式、可视化的呈现方式提供给用户，从而帮助用户实现有效、专业的数据分析。

● 2027年，数据分析与可视化领域的重大进展或热门话题将是什么？

2027年，数据分析与可视化领域的进展将使不同语义层面的跨数据源和跨领域的分析成为可能。广大用户将能够利用自然语言等更方便的交互机制，快速有效地从数据中获得与业务和生活各方面相关的广泛而深入的信息。

孙晓艳 微软亚洲研究院主管研究员

● 2017年，计算机视觉领域将有哪些重大进展？

2017年的重大进展就是：基于深度学习的计算机视觉技术将在2017年继续高速发展。这将在接近人类能力的高精度对象识别、轻量级便携式视觉系统以及视觉平台的广泛应用中充分体现。

● 2027年，计算机视觉领域的重大进展或热门话题将是什么？

到2027年，视觉"眼"将无处不在。我们将拥有高度发达的成像设备、强大的计算资源以及整合深度和广度学习的视觉技术。这些技术的进步使得无处不在的视觉"眼"能够"看到"并在日常生活和各种职业中为人们赋予更大能力——从制造和医疗一直到金融和安全。

Jennifer Chayes 微软新英格兰研究院院长兼微软纽约研究院院长，杰出科学家

● 2017年，机器学习算法领域将有哪些重大进展？

深度学习正在许多方面改变我们的技术，然而，当今的深度学习算法很大程度上仍然是启发式的——基于该领域领导人物的经验和直觉。2017年，我们将开发一种更有原则性的方式来理解深度学习，以及由此得出更加强大的算法。这些洞见将来自诸多领域，包括统计物理学和计算机科学的交叉学科。

● 2027年机器学习算法领域的重大进展或热门话题将会是什么？

人工智能和机器学习算法大大改善了我们的生活。然而，目前的算法常常再现了我们数据中的区分性和不公平，并且受到输入误导数据的操纵。未来十年算法方面的一个重大进步将是开发出更加公平、可追责、且更不容易被操纵的稳健算法。

Susan Dumais 微软雷德蒙研究院副院长，杰出科学家

- 2017年，搜索和信息检索领域将有哪些重大进步？

搜索和信息检索中的深度学习将盛行。过去几年，我们看到语音识别、图像理解和自然语言处理等领域取得突破，这是全新深度学习架构与更多数据和更强大计算能力结合的结果。2017年，深度学习模型将继续提高网络搜索结果的质量，实现文档理解和查询契合度方面更全面的改进。

- 2027年，搜索和信息检索领域的重大进步或热门话题将是什么？

搜索框将消失。它将被更加无处不在、嵌入式、对上下文敏感的搜索功能所取代。我们通过语音查询而看到这一转变的开始——尤其是移动和智能家庭场景。这一趋势将会加速发展，并且能够发起包含声音、图片或视频的查询，无需提出明确的查询就能利用上下文来主动检索与当前位置、内容、实体或活动相关的信息。

Kalika Bali 微软印度研究院研究员

- 2017年语音和自然语言处理领域将有哪些重大进展？

2017年，我们的语音和语言处理技术应用将会涉及越来越多的语种。这并不仅仅意味着我们将在我们的产品上添加更多语言。我们确实会这样做，但是我们还将开发能够理解、处理并生成语言的系统，可供有英语-西班牙语、法语-阿拉伯语或印度语-英语需求的用户使用，使她可以从一种语言轻松切换到另一种语言——即使在同一场对话、聊天，甚至是在同一个句子中也能轻松切换。

- 2027年，语音和自然语言处理领域的重大进展或者热门话题将会是什么？

语言模型将在认知模型中确立牢固的地位，让人工智能（AI）系统能够相对轻松地进行推理，并与人类沟通，灵活转换并适应不同的社交场合、谈判、辩论和说服，就像人类一直以来一样。计算社会语言学和语用模型将在创造拥有社会文化意识的人工智能中发挥重大作用。

Mar Gonzalez Franco 微软研究院新体验与新技术部研究员

- 2017年，虚拟现实领域将有哪些重大进展？

2017年，我们将看到拥有更好的人体追踪技术的虚拟现实设备出现。其积极的结果就是能够从第一人称的视角体验到虚拟化身。

- 2027年，虚拟现实领域的重大进展或热门话题将是什么？

到2027年，虚拟现实系统将无处不在，并且能够提供丰富的多种感官的体验，从而产生融合或改变"感知现实"的幻觉。利用这一技术，人类将重新训练、调整并改进其感知系统。与目前只能刺激视觉和听觉感官的虚拟现实系统相比，未来虚拟现实的体验将扩展到其它感官——其中包括通过触觉设备而获得的触觉。

阅读全部十七位女性研究员观点
扫描二维码即可查看

每一个人的人工智能

作者：微软全球执行副总裁、微软人工智能及微软研究事业部负责人 沈向洋

编者按：

1月14日，未来论坛2017年会暨首届未来科学颁奖典礼在北京拉开帷幕，微软全球执行副总裁、微软人工智能及微软研究事业部负责人沈向洋受邀在此次论坛上发表了主题演讲。本文包括沈向洋博士此次演讲内容的文字整理，希望与大家分享微软对人工智能的理解和对未来的美好愿景。

人类今天面临的最重要课题之一就是人工智能。纵观人类历史上的诸多创新，像汽车、印刷机、蒸汽机、载人飞机、火箭和人造卫星……它们都不只是孤立的创新，更是推动人类在浩瀚无垠的各领域里迅猛发展的加速器。

人工智能也许是迄今为止最强有力的加速器，因为它不仅使人类的能力得以延展，还让我们的脑力变得更加强大。它不仅仅是数字时代的一个进步，而是包括图灵在内的很多人，在电脑诞生前就梦想到的终极目标。大型机、个人电脑、互联网、移动电话、云……这一切都是人工智能的前奏。

回溯过去几十年，人工智能的发展历经高潮与低谷。现在又有哪些不同？三大创新最终将人工智能带上了快车道。互联网、大数据和传感器随处可见；特别是基于云平台的大规模计算能力；得益于突破性算法的发展，计算机已经可以凭借深度学习独立完成更为复杂的任务。

正因为这些突破，人工智能已经无处不在——它可以为你推荐所选，保护你免受信用卡诈骗，个性化处理你的新闻推送，并让你的家智能互联。微软正在保护你的Outlook邮箱不受垃圾邮件骚扰，并赋予智能聊天机器人小冰以声音与个性。然而即便如此，我们现在不过是触碰到了人工智能潜力的冰山一角。

对于微软而言，人工智能不只是一个关注的焦点，更是我的第一要务。过去25年来，我们一直在通过各种方式为这一轮创新做准备。25年前比尔·盖茨创建微软研究院时，他就曾设想有一天创造出通用人工智能，一个会看、会听、能讲，并且能够理解人类思维的电脑。

多年来，我们吸引了全球的顶尖人才加入我们的研究院，一起迎接这些挑战。我们的国际团队是首屈一指的。锦上添花的是，我们刚刚宣布收购Maluuba，这是一家总部位于加拿大蒙特利尔的公司，拥有全球最领先的人工智能研究和工程团队。

除此之外，作为全球最重要的深度学习研究专家之一、蒙特利尔学习算法研究所所长和Maluuba顾问，Yoshua Bengio也将成为微软顾问，并直接和我互动。

我的团队已经取得了一些激动人心的突破。就计算机视觉而言，我们的视觉算法正接近人类的水平！微软在诸如ImageNet这样的图像识别比赛中屡次斩获大奖。更令人兴奋的是，我们在视觉算法方面的成果已经被用在微软自己的产品上，比如HoloLens；同时也运用在日常应用上，比如为UBER这样的客户验证驾驶员身份。

计算机正在学习像我们人类一样去看，并为它们识别出的东西命名。现在它们也能像我们人类一样倾听和识别语音。三个月前，微软的语音识别技术在行业标准Switchboard测试中首次达到了人类同等水平，而这个目标在几年前还看似遥不可及。

随着我们在语音识别领域能力的提高，我们在面对诸如自然语言等更复杂的挑战上也取得了实质性进展。世界上有6500多种语言。试想如果我们都能相互理解，这将意味着什么！就在上个月，我们的Microsoft Translator取得了巨大的飞跃。Microsoft Translator能实现包括中文普通话在内9种不同语言之间的实时语音翻译，最多允许100人同时进行群组交流。我们采用了一项令人兴奋的深度强化学习技术，并在此基础上运用了对偶学习，来加速提高翻译能力。

我们正在接近目标，让计算机可以做某些人类力所能及的事：看、听、并开始描述它们周围的世界。但人类更了不起的地方在于，我们不会止于感知，我们会继续向认知深入。我们会将想法、概念与世界上的人、地点与事物联系起来……并了解它们彼此的关系。这就是我们所说的“常识”。

这对人类来说再普通不过，但对计算机来说却是一项异常艰巨的任务。多年来，源于在必应搜索（Bing）的研究和实践，我们建立了一个庞大的知识图谱。我们了解数十亿计的实体——人、地点、事物。我们既拥有“关于世界的知识”、也拥有令人叹为观止的“关于工作的知识。”以Office 365和LinkedIn为例，通过将这些特有的工作数据与视觉、语音以及语言识别能力相结合，微软就可以打造出能将个人生活与职场生活统一相连的人工智能助理。

虽然取得了这些进展，但人工智能还未能实现通用智能。计算机可以顺利完成特定任务，但当面临通用任务时，它还不如一个小孩做得好。我们在人工智能涉及的所有领域里都已取得了巨大的进步，但要实现通用的人工智能，仍有很多工作要做。

首先，我们要让计算机理解我们，而不是强迫人们去理解计算机。实现这一想法的核心在于双向对话：计算机很难像“人类”一样讲话。我们称之为“对话式人工智能”，这也是人机交互下一步要实现的目标。

回溯图形用户界面的诞生，它让每个人都能使用个人电脑。而对话式人工智能更是颠覆性的。我们能在任何地方使用它，不再受屏幕与键盘的束缚。我们也不必再去适应计算机的规则。相反，计算机会来理解我们。当计算机能像“人类”一样说话时，每个人都可以和计算机对话与协作。

为实现这个目标，微软着眼于“人类智能”的含义。人们明白许多事是没法用IQ（智商）来衡量的。在你我交谈的时，我们时时刻刻都会留意对方的语调和手势。我们能理解文字游戏——比如幽默、讽刺、嘲笑，或夸张。我们称这种情绪理解能力为EQ（情商）。要打造成功的人工智能，EQ与IQ同样重要。我们要让人工智能助理理解我们真正的意图，就像你我一样。

拥有EQ和IQ，智能助理能让计算实现真正的人性化。它们将了解我们：认识我们是谁，甚至明白我们的感受。它们将预判我们的需求，让我们按计划行事。它们不会拘泥于任何一种单一的设备；相反，只要我们需要，它们能随时陪伴左右。比如目前，我们已经与汽车厂商合作，赋予小娜这样的能力。

而最终，得益于人工智能，计算机会越来越像人类。但我想人工智能也许还可以让人类变得更好：人工智能会增强我们人类的能力。它会赋予每个人以“超能力”来应对各种挑战——包括最艰巨的挑战。

以教育与英语学习为例。要掌握英语口语，每个学生都需要跟老师一对一练习，这样才能直接得到老师的反馈。可老师没有“分身术”——他们可以同时指导几十个学生，但给每个人的关注也因此降低。人工智能就能解决这个问题。微软研发了自己的虚拟英语老师——微软小英。基于微信服务号，微软小英能为用户提供免费服务。现在，微软小英已拥有超过80,000名用户，我们正与中国日报双语新闻微信公众号合作，读者可以用它测试自己的英语词汇量。

这个世界充满挑战。但我们非常乐观！借助人工智能，我们可以在人类力所能及、想象力所及的范围内取得更迅猛、更深入的发展。人工智能的故事还没讲完，正如计算机领域的传奇人物Alan Kay所说：“预测未来最好的方式就是创造未来。”

作为与人工智能共同生活的第一代人，这就是我们要做的事情。在世界各地，微软“梦之队”正在努力践行我们的使命——予力全球每一人、每一组织，成就不凡。

我们正在开发的人工智能，将增强我们作为人类所独有的素质——我们的好奇心、爱心、创造力，以及我们对于一个更美好世界的共同渴望。

如何做好计算机视觉的研究？

作者：微软亚洲研究院资深研究员 华刚

简单说一下这篇文章的背景：从我去年（2015年）回到微软亚洲研究院之后接触到很多聪明的实习生。一方面感受到他们对计算机视觉研究的热情，另一方面也有感于他们对计算机视觉研究认知的局限性，或者说大一点，是基本研究方法和思路上的局限性，就有想法要对如何做好计算机视觉的研究写点什么，但一直也没有找到合适的机会。最近计算机视觉领域国际权威、加州大学洛杉矶分校的朱松纯老师发表了一篇关于计算视觉的三个起源和人工智能的评论，引起了很大的反响。朱松纯老师的评论全面深刻，我想借着这个机会，结合朱老师评论的内容和我在计算及视觉领域15年的研究经历，也来谈谈如何做好计算机视觉的研究，希望对领域内的学生和年青的研究员能有所帮助。

“如何做好计算机视觉的研究？”

要回答这个问题，我们先要对这个问题的关键词进行分析：如果去掉“计算机视觉”这个限定词，这问题就变成了“如何做好研究？”那么，要回答这个问题，我们就要知道“什么是好的研究？”而要定义什么是好的研究，必须回到根本，先要知道“什么是研究？”

我们的讨论就从这个问题开始。

● 什么是研究？

一个被普遍接受的对研究的广义定义为：研究是为了产生新的知识或者是为已有的知识设计新的应用的系统性的工作。因为我们今天的讨论其实更多集中在科学研究上，先确定狭义的研究的定义为：利用科学的方法来调查解释一个现象或者获取新的知识。

综合这两个定义，可以看到科学研究从本质上是由三个基本的要素构成：1)目的：产生新的知识或者是设计出新的应用；2）手段：科学的方法。缺少这两个要素任何之一都不构成科学研究; 3) 成果：新的知识。所谓新的知识，必须是前人不知道的东西。

我们很多同学和年轻的研究员认为研究就是写论文、研究成果就是论文，这其实是在观念上走进了一个误区。论文是系统阐述新的知识、新的应用，以及阐述获取这个新知识或者新应用用到了什么样的科学方法的一个载体。论文，作为阐述研究成果的主要手段，必须经过同行的评议通过才能正式发表和被认可。

在人工智能进入第三个热潮之际，我们看到各种各样关于AI的媒体报道层出不穷，一方面，这对大众普及了AI各方面的知识，是积极的。但从另一个方面讲，很多观点没有经过仔细的推敲，也没有同行的评议，一些谬误或者是夸大的观点可能因为广泛传播而被大众接受，结果产生负面的社会影响。这就提醒我们相关领域的研究人员，在对大众媒体去做一些评论的时候，必须仔细斟酌，尽量不传播没有得到检验的观点。

这就谈到第二个问题：

● 什么是好的研究？

不同领域的研究员对这个问题可能会有不同的看法。

从计算机科学的角度来讲，尤其是计算机视觉的研究，无论是理论的还是实践的，我们的研究成果最终是要解决现实世界的问题的。在这个方面，我印象比较深刻的还是我在西安交通大学读研究生的时候，沈向洋博士2001年在西安交大做报告提到的一个观点：最好的研究员发现新问题；好的研究员创造新方法解好问题；一般的研究员跟随别人的方法解问题——大家在多次这里看到“新”这个关键词，创新是研究的本质。

有了这些铺垫，我们首先定义什么是最好的研究。通常认为一个领域中对于某一个问题最好的研究工作有三种：第一篇论文 (The First Paper)，最好的一篇论文 (The Best Paper)，以及最后一篇论文(The Last Paper)。这第一篇论文的含义是说这篇论文率先提出了一个好的问题和方向。最好的一篇论文是什么？那一定是开创性地提出了一种解法，启发了最终解决这个问题的途径。至于最后一篇论文，那一定是彻底把这个问题解决了，

从此以后这个问题不再需要继续做进一步的研究。

从计算机视觉领域举一个具体的例子来讲，Harris Corner Detector属于最早的一批在图像中检测角点的论文，可以归为第一篇之列。David Lowe博士的SIFT特征检测和局部描述子，可以归为在这个方向上最好的论文之列。那么这个方向的最后一篇呢？我认为可能还没有出现。具体到我自己的研究工作，在局部描述子这个方向上，我跟我的同事Matthew Brown和Simon Winder在2007年到2009年之间所做的一系列用机器学习的方法来建立描述子的工作，也实际上为提高局部描述子的性能提供了一个新的思路和方法。

对于我们很多研究员和学生来讲，一辈子可能都做不到这三种最好的研究工作之一。那是不是就等于说你不能做好的研究工作或者根本不用考虑做研究了呢？肯定不是这样。科学研究是一个共同体。这些最好的研究工作也是在前面很多很多非常扎实（solid）的研究工作的基础上发展出来的。因此，对于年青的研究员和学生而言，应该胸怀大志，去追求做最好的研究工作，但从实际执行上来讲，还是要把一项一项具体的工作先做扎实了。

怎么做到把研究工作做扎实了？首先，你必须对你要解的问题有一个全面深刻的了解，包括为什么要解这个问题、解这个问题有什么意义呢、以前有没有试图解决同样或者类似问题的先例，如果有，你就要全面了解前人都提出了什么样的解法、他们的解法都有什么样的优势和缺陷……最后，你的解法解决了前面这些解法不能解决的问题呢，或者是你的解法处理了什么样的他们不能处理的缺陷了？这些问题的答案如果都有了，那么，在写论文的过程中要注意的就是：1）你的假设是什么？2）你怎么验证了你的假设？这个验证既可以是理论上的证明，也可以是实验的验证。我们很多学生和年青的研究员，写论文的时候没有找到内在的逻辑关系，很多观点都是似是而非。或者说重一点，在论文撰写方面的训练严重不足。你的研究如果到了写论文的阶段，那就必须要有明确的观点提出来。这个观点必须明确无误，只有这样你才能被称为形成了新的知识。你的每一个观点都必须在理论上或者是实验中得到验证。另外，论文的撰写是为了让人看懂，不是让人看不懂，所以我们在撰写过程中必须尽量保证不去假设读者已经拥有了某些方面的知识。做好了这些，基本上你就有很大的可能性能够做出扎实（solid）的研究工作。

然后回到我们讨论的主题：

● 如何做好计算机视觉的研究工作？

其实，要回答这个问题，将我上面讲的所有观点加上“计算机视觉领域”这个限定词就行了。我这儿结合计算机视觉研究的一些现状及朱松纯老师的一些观点来进一步谈谈我的观点。

首先谈谈我观察到的一些现象。很多年轻的学生，现在讨论问题的时候都用这样的谈话：我发现用FC6层的特征，比用FC7层的特征，在某个图像数据集上比现在最好的算法提高了1.5%的识别精度，老师我们可以写论文了（如果大家不能理解这句话，FC6和FC7是表示AlexNet的两个中间输出层）。我想请问，你在这个过程中发现了什么样的普适的新的知识吗，又或者，在不是普适的情况下，你在什么限定条件下一定能够看到这样的识别精度提高了？

不错，提高识别精度是一个很好的目标，但要注意，计算机视觉的研究是要解决识别的问题，不是解某一个图像数据集。这些图像数据集提供了很好的验证你的假设和方法的手段，但如果你没有遵循科学的方法和手段去设计你的算法和实验，你也不可能得到一个科学的结论，从而也不能产生新的知识，更不用谈对这个领域做出贡献。朱松纯老师在他的评论中提到，很多学生认为，计算机视觉现在就是调深度神经网络的参数，也就是说的这个问题。

所以，具体到对于刚开始从事计算机视觉研究的学生来讲，要做好这方面的研究，我觉得第一步还是要系统学习一下计算机视觉的课程，全面了解一下计算机视觉这个领域的来龙去脉、这个领域都有哪些基本的问题、哪些问题已经解得比较成熟而哪些问题还在初级阶段……这里，推荐所

有的学生学习两本经典教材《Computer Vision: A Modern Approach》和《Computer Vision: Algorithms and Applications》，可以先读完第一本再读第二本。

只有对这个领域有了一个初步的全面了解，你才能够找到自己感兴趣的那个问题。在众多的问题当中，你是希望做三维重建，还是做图像识别、物体跟踪，又或是做计算摄影呢？做研究其实不是一个完全享乐的的过程，你必须要有足够的兴趣来保证你能持续地走下去，这在你感觉自己当前研究的思路走不下去的时候尤其具有重要意义。当你确定你感兴趣的问题，你应该首先全面调研一下这个问题的来龙去脉。这就意味着你不能只读过去五年的论文。你可以从过去一年的论文开始，慢慢追溯回到过去很久的相关的论文。有些时候，你会惊讶地发现前人想问题的深度。研究的英文单词是Research，拆开是Re-Search，用中文直译就是重新搜索和发现，而不是直接发现，其实就是说你要首先对这个问题做追本溯源。朱松纯老师提到的我们很多学生现在不读五年以前的论文，说的也是这个道理。

当你做好了这些，你必须钻进计算视觉的一个小的领域。人的精力是有限的，这就意味着你不可能把很多事情同时做好，所以在你选好方向之后，就要把你的精力集中在你感兴趣的一个问题上， 努力成为这个方面的专家。研究是一项长跑，很多时候，你在一个方向上比别人坚持久一点，你就有机会超越他而成为某个方面的专家。

最后，我也来谈谈深度学习对计算机视觉的影响。在这里，我对马里兰大学Rama Chellapa教授在Tom Huang教授80岁生日论坛上表达的观点非常认可，他认为，深度学习网络就像一个Pasta Machine：你把该放的东西放进去，它能给你产生好吃的Pasta。同时它也是一个Equalizer：无论你在计算机视觉领域有40年的经验还是0年的经验，只要你会用Caffee，你在一些问题，比方说图像识别上，都能产生差不多的结果。他开玩笑说这有点伤自尊（It hurts my ego!），但我们还是应该把它作为一个好的工具拥抱它。我想，他的言外之意，是我们的研究应该做得更深，要去理解这个工具为什么能够工作得比较好，从而产生新的知识去指导将来的研究和应用。

我认为，对于年轻的学生来讲，从深度学习的方法开始学习没有什么问题，但必须要进一步去了解一下其他的数学和算法工具，像统计贝叶斯的方法、优化的方法、信号处理的方法等等的。计算机视觉的问题，其本质是不适定的反问题，解这一类问题需要多种方法的结合。这里面有深度学习解得比较好的问题，像图像识别，也有深度学习解不了的问题，像三维重建和识别。

任何研究领域包括计算机视觉的研究，对处在研究初期的学生而言，更重要的是掌握足够的数学工具，培养一种正式思维（Formal Thinking）的能力，这样，遇到实际的问题就能以一种理论上正确的思路去解决这个问题。

作为结束语，我想对在从事或者有志于从事计算机视觉研究的学生说，计算机视觉的研究处在一个非常好的时期，有很多我们原来解不了的问题现在能够解得比较好了，像人脸识别，尽管我们其实还没有从真正意义上达到人类视觉系统对人脸识别的鲁棒程度。但我们离真正让计算机能够像人看和感知这个世界还有很远的距离。在我们达到这个目标之前，深度学习的方法可能是这个过程中一个重要的垫脚石，同时我们还要将更多的新的方法和工具带入这个领域来进一步推动这个领域的发展。

了解微软亚洲研究院
视觉计算组
扫描二维码即可查看

微软亚洲研究院四位研究员获选IEEE Fellow和ACM Distinguished Member

2016年末，IEEE（国际电气电子工程师学会）和ACM（美国计算机协会）先后公布了2017年度的院士名单（IEEE Fellow）和2016年度杰出会员名单（ACM Distinguished Member）。其中，微软亚洲研究院共有四名研究员获选，他们分别是：刘铁岩博士获得IEEE Fellow和ACM Distinguished Member双重荣誉，华刚博士、梅涛博士以及郑宇博士荣获ACM Distinguished Member称号。这一殊荣代表着计算机学术界对微软亚洲研究院四位研究员在各自领域所取得的卓越成就的高度肯定。正是这些优秀科学家对科研的执着、信仰和不断创新，推动着科学技术的持续发展，潜移默化地改善着我们的生活。

让我们共同祝贺四位研究员获此殊荣！

• 刘铁岩博士，微软亚洲研究院首席研究员

刘铁岩博士因对机器学习方法及其在网络搜索和在线广告领域应用做出的杰出贡献，同时获选国际电气电子工程师学会院士（IEEE Fellow）以及ACM杰出会员（ACM Distinguished Member）。

刘铁岩博士，现任微软亚洲研究院首席研究员，领导机器学习和人工智能方面的研究工作。他的先锋性工作促进了机器学习与信息检索之间的融合，被国际学术界公认为“排序学习”领域的代表人物。刘博士发明了一系列具有高度影响力的算法和理论，包括列表级排序算法（Listwise ranking）、BrowseRank、关系排序、以及统计排序理论等等。他在该领域的学术论文已被引用近万次，并受斯普林格出版社之邀撰写了该领域的首部学术专著（并成为斯普林格计算机领域华人作者的十大畅销书）。刘博士的工作推动了排序学习成为一个重要的学术分支，并至今仍然是信息检索领域主流学术会议的核心研究方向之一。

同时刘铁岩博士在分布式机器学习方面也颇有建树。早在2005年，刘博士就开发了当时世界上最大规模的文本分类器（text classifier），它可以高效地完成超过25万个类别的分类任务。近年来，他又领导开发了包括LightLDA（目前世界上最大的主题模型，可以处理100万个主题）和LightGBM（目前世界上最高效、扩展性最好的GBDT算法）在内的大规模机器学习算法，并通过微软分布式机器学习工具包（DMTK）开源，受到了广泛关注。

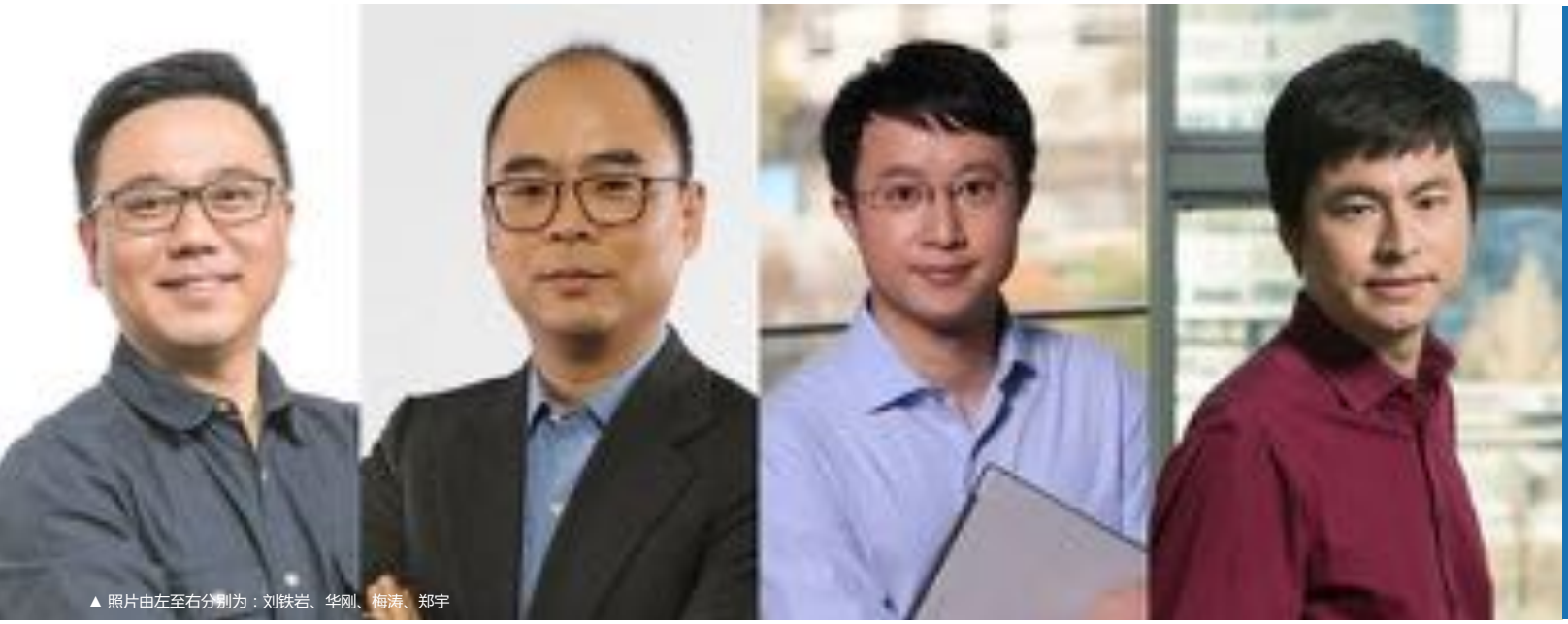
刘铁岩博士在顶级学术会议和期刊上发表了数百篇论文，并多次获得最佳论文奖、最高引用论文奖、研究突破奖。他曾受邀担任了包括WWW、KDD、NIPS、AAAI在内的顶级国际会议的组委会主席、程序委员会主席、或领域主席；以及多个国际期刊副主编。除此之外，他还在很多学术组织中扮演着重要角色，包括中国计算机学会（CCF）学术工委，中文信息学会信息检索专委会副主任等。正是因为刘铁岩博士的杰出学术成就和社会影响力，他被聘任为美国卡内基·梅隆大学（CMU）客座教授、英国诺丁汉大学荣誉教授、以及中国科技大学、中山大学、南开大学博士生导师。

• 华刚博士，微软亚洲研究院资深研究员

因对多媒体和计算机视觉研究领域做出的杰出贡献，华刚博士获选ACM杰出会员（ACM Distinguished Member）。

华刚博士是微软亚洲研究院资深研究员，现任微软亚洲研究院计算视觉组负责人。他的研究重点是计算机视觉、模式识别、机器学习、人工智能，和机器人，及相关技术在云和移动智能领域的创新应用。华刚博士是通过对语境建模来使用弱监督或者无监督方法来解决无限制环境下计算机视觉问题的倡导者。其研究在学术界和工业界都产生了广泛的影响。华刚博士在2008年创立和奠基了微软的人脸识别引擎。到现在已发展成为微软认知服务（Cognitive Services）中的人脸识别应用程序接口（Face API）。

华刚博士于2011年在国际顶级期刊IEEE Trans. on Pattern Analysis and Machine Intelligence（IEEE模式分析和机器智能汇刊）上领导组织的“现实世界人脸识别”专刊对推动无限制环境下人脸识别的研究产生了深远的影响。因其在图像和视频中无限制环境人脸识别研究所做出的杰出



▲ 照片由左至右分别为：刘铁岩、华刚、梅涛、郑宇

贡献，2015年华刚博士被国际模式识别联合会（International Association on Pattern Recognition）授予“生物特征识别杰出青年研究员”。

华刚博士已在国际顶级会议和期刊上发表了120多篇同行评审论文。他将担任CVPR 2019的程序主席，以及CVPR 2017和ACM MM 2017的领域主席。华刚博士曾担任十多个顶级学术会议的领域主席，以及IEEE Trans. on Image Processing（2010-2014）的编委。目前，华刚博士还担任着IEEE Trans. on Image Processing、IEEE Trans. onCircuits Systems and Video Technologies、IEEE Multimedia、CVIU、MVA和VCJ的编委。

● 梅涛博士，微软亚洲研究院资深研究员

因在大规模视频分析领域做出了杰出贡献，梅涛博士获选ACM杰出会员（ACM Distinguished Member）。

梅涛博士是微软亚洲研究院多媒体搜索与挖掘组资深研究员。他的主要研究兴趣为多媒体分析和计算机视觉，现已发表论文100余篇，先后10次荣获最佳论文奖，拥有17项美国专利，其研究成果先后十余次被转化到微软的各产品和服务中，如Office、Windows Photo、Bing、OneDrive、微软小冰等。

梅涛博士是视频内容理解和搜索的先锋者。在视频理解领域，他率先提出对视频底层视觉表示的上下文与高层语义之间的相关性分别进行建模，他还开创性的提出了基于内容搜索的视频标注方法，其相关研究先后两度荣获多媒体顶级期刊IEEE Trans. on Circuits and Systems for Video Technology（TCSVT）和IEEE Trans.on Multimedia（TMM）的年度最佳论文奖。梅涛博士的另一项创新性工作是荣获ACM多媒体年会最佳论文奖的“视觉查询推荐”，该技术可在大规模视觉搜索系统中通过提供结合文本和图像的联合查询推荐来实现高效搜索。近年来，梅涛博士还领导了视频和自然语言（Video and Language）的研究，并已将相关研究成果成功的应用到微软聊天机器人小冰的视觉聊天中。

目前，梅涛博士同时担任IEEE和ACM两大多媒体顶级会刊TMM和TOMM（ACM Trans on Multimedia Computing, Communications, and Applications）的编委，他还是多个国际多媒体会议的大会主席和程序委员会主席，以及十多个国际会议的领域主席。他还同时担任中国科学技术大学和中山大学的博士生导师，以及ACM SIGMM中国分会的大会主席。

● 郑宇博士，微软亚洲研究院资深研究员

因推进城市计算和时空数据挖掘领域的科学研究，郑宇博士获选ACM杰出会员（ACM Distinguished Member）。

郑宇博士，微软亚洲研究院资深研究员，他的研究兴趣包括大数据分析，时空数据挖掘，人工智能和普适计算。郑宇博士是城市计算领域的先驱，他的愿景是以连接传感技术、数据管理和分析模型以及可视化方法，创造出改善城市环境、提高人们生活质量和城市运行系统的三赢解决方案。他也推进了城市计算中最常用的数据形式之一——“轨迹数据”的数据挖掘技术。

同时，郑宇博士还提出了数据融合方法，以发挥不同领域的多个时空数据集的知识力量。他的跨域数据融合方法可以通过融合其他领域的知识（如交通和气象学），来解决一个领域的挑战（例如空气质量），使人们能够利用稀疏和噪声数据来解决城市计算中的问题。

2013年，郑宇博士因在城市计算领域的贡献被MIT科技评论评为全球杰出青年创新者（MIT TR35），并作为现代创新者代表登上了美国《时代》周刊（Time Magazine）。

目前，郑宇博士是人工智能顶尖国际期刊（SCI一区）ACM Transactions on Intelligent Systems and Technology（ACM TIST）的主编以及IEEE旗舰刊物 IEEE Spectrum的顾问委员会成员。他还是IEEETransactions on Big Data和GeoInformatica的编委，以及SIGKDD中国分会创始秘书。

这里是中国人工智能的摇篮，他们的故事仍在继续

作者：李北辰 《界面》

2016年11月14日上午9点，中关村丹棱街5号微软大厦，李开复、张亚勤、沈向洋、洪小文、王坚等中国IT界的技术大佬，以及曾在微软亚洲研究院工作过的近200位院友“同框出现”，宣布“微软亚洲研究院院友会”正式成立：沈向洋担任会长，李开复和张亚勤任名誉会长，他们都是这家研究院曾经的院长。

坦率地讲，夹杂在繁芜的科技资讯信息流之中，这很难称得上是一条惹眼新闻，至少相比上述大佬的同框，几天之后张亚勤和沈向洋出现在乌镇的那一张饭局合影要更为被媒体所追捧。

但科技记者无法忽视的是，多年来，在稿子中描述圈内大佬过往履历时，“微软亚洲研究院”是一个高频词汇。翻看人物图谱：这里是阿里云之父、前金山CEO、百度总裁、小米总裁、海尔CTO、联想CTO……互联网高管、以及浙江大学、中国科学技术大学、人民大学等国内一线大学计算机/信息技术院系负责人的上一处居心之所；这里和跨国巨头在中国的“办事处”关系不大，它更像是一座拥有18年校龄的“大学”。

中国IT界的黄埔军校，有人喜欢这么说。在院友会成立仪式上，院友代表王坚感喟：没有人可以绕开这个机构去谈论过去十几年的中国科技史。

一切得从往事谈起。

● 九八年的召唤

九十年代初——在最不可一世的岁月，来自西雅图的软件巨人却为创新乏力所困。微软前任首席技术官麦尔伏德在1991年建议盖茨成立带有“军备竞赛”属性的创新研究院，进行基础科学研究，以主动迎合技术世界的不确定性。这年，微软首座研究院在雷德蒙成立。

六年之后，麦尔伏德送给了盖茨第二个建议：将研究院开到中国。理由颇为机敏，到1990年代为止，美国一直是这个世界上每年诞生计算机博士最多的地方——直至被中国超越，这被麦尔伏德视作中国IT业态苏醒的前奏。

1998年，微软中国研究院在北京成立。37岁的语音识别专家李开复博士成为首任院长。18年后已成谈资的插曲是：当得知李开复要以“研究院”作为这座机构的后缀时，沈向洋博士大笑，“叫个‘所’都大了”，2001年中国研究院升级为微软亚洲研究院则是后事了。

▲ 微软中国研究院建院合影 摄于1998年11月

至少在1998年，沈向洋有充分的理由自嘲。现任院长洪小文后来回忆：“研究院建立前，我到北京参与最初一批员工招募。有些候选人是国内高校博士生，我们只能通过固定电话联络，那时学校里通常一幢宿舍大楼只有一部电话，经常需要麻烦宿管老师或同学帮着叫人来接听。没有几个网站，没有太多信息资源，因而沟通时需要花很多时间同步基础信息，效率比较低。Email倒是能用，但拨号上网资费昂贵又不稳定，实时在线不大可能。”

1998年，每小时8块钱的上网费和4块钱的电话费让上网属于奢侈消费；这一年，配备Windows 95的个人电脑也实属高大上的新锐电器；在

▲ 照片由左至右分别为：洪小文、张亚勤、李开复、沈向洋、王坚

更为宏大的叙事里，这一年，中国GDP刚突破万亿美元大关，IT产业所占比重仅约为4.8%。

然而同样在1998年，互联网作为社会基础设施迎来破茧之日，一批互联网明日巨星集体相约九八：1998年9月4日，谷歌正式创立；中国则更是普遍把这一年视作商业互联网元年：2月，张朝阳创办搜狐；6月，刘强东成立京东；11月，腾讯诞生；12月，新浪上线。

机遇与挑战在1998年显得尤为冲突——但考虑到成立研究院的愿景，天平似乎更偏向挑战一端。

早在筹备阶段，比尔盖茨和里克·雷斯特博士——微软全球研究院体系的缔造者，就为研究院确立了三条宗旨：推动整个计算机科学前沿技术发展；将最新研究成果快速转化到微软关键产品之中，以帮助用户改善计算体验；着眼于下一代革命性技术研究，助力微软实现长远发展战略和对未来计算的美好构想。

所以你可以想象李开复说服海外精英归国的不易，上述宏愿与当时中国的外部环境格格不入。不过事实美好的一端也成为18年后院友会成立仪式上李开复追忆的画面：“我在卡内基梅隆大学读书时有位同学叫沈为民，写代码非常厉害，我说你是怎么做到的。他说我们在国内都是在纸上写代码，老师则用脑子来运行这个程序。当时给我的感触是，如果能在纸上培养出沈为民这样的人，这样的人才肯定还有很多，一起做研究院一定能做出来。我当时也是抱着这样乐观的心态就回来了。”而中国的人才也是当年比尔盖茨所看重的。

后面的事情令人欣喜，研究院开始在全球顶级学术会议与期刊崭露头角，尤其2002年，他们在计算机科研领域最富盛名的国际图形学年会SIGGRAPH上发表了4篇论文——长久以来这种国际大会鲜有华人身影。

有人后来回忆，当年前去知春路希格玛大厦（研究院旧址）“朝圣”的年轻人一定会颇为失望，一切看上去与其他企业别无二致，除了目光中的这位“理工男”可能刚在国际期刊发表了一篇重要论文。

● 流动的“黄埔军校”

如今十八年过去，微软亚洲研究院当年的愿景已兑换成国际顶级学术会议与期刊上的4000余篇论文。更重要的是，脱胎于这里的技术也转移到微软产品矩阵中——Office、Windows、Bing、Kinect以及小冰、Cortana、Skype Translator等人工智能产品，都有亚洲研究院的身影。

事实上，在人工智能还是个模糊概念，甚至不被学术界普遍承认的岁月，研究院就已将其视作面向未来的基础研究。1999年研究院成立一周年时即确立了最初的几个研究中心：语音组，多媒体组，图形图像组——如今它们都更像是人工智能概念的一部分。

李开复多年后坦言：中国如今能在人工智能领域扮演重要作用，很大程度上要归功于研究院很早开始便在语音识别，自然语言理解等众多人工智能核心技术进行研究。“中国今天的成就真的应该感谢微软亚洲研究院。”

不过，这种“感谢”对于研究院来说，似乎也意味着成长的烦恼。几年前开始，中国科技公司就积极网罗人工智能领域的精英，可以想象，微软亚洲研究院作为最密集的人才聚集地成为了被挖的目标。

事实上，研究院早已熟悉了人才的流动——这也是成立院友会的原因之一。

十八年前，这个国度最顶尖的技术头脑纷纷从美国的高校和科技企业来到微软；十八年后，技术后辈有了更多良禽择木的可能——聪明的大脑不再浓烈地聚集一地，而是稀释在各个地方。他们有些人继续留下来从事自己最热爱的研究，有些人选择离开去领导更大的团队做产品开发，有些人选择到大学任教，有些人则选择在中国这片创业热土圆梦。

如今微软亚洲研究院全球院友已超过5000名，其中有150多位院友在世界各地顶尖高校执教，15位院友是电气电子工程师学会（IEEE）院士，美国计算机协会（ACM）院士或美国人工智能学会（AAAI）院士，20位院友入选了中国千人计划和国家杰出青年科学家基金；超过10位院友在互联网企业担任CEO，CTO或CIO，范围几乎覆盖所有中国最顶尖的互联网企业；80多位院友在创业大潮中有所作为，他们有些是创始人，有些是技术领导者；而从微软亚洲研究院走出的实习生也有超过500名活跃在中国IT产业的各大企业。

● 人工智能的延续

命运自有时。现在研究院的初心倒是与十八年前没太大不同。洪小文在一次院友聚会中有些动情地说到：“有些人离开了几年了，有些人刚刚离开，我可以保证的是，我们这群人还会好好地在这边专心致志地做科研创新，我们不是产品部门，我们也不是学校，但是我们就是这样子的一个扎实的单位，研究院存在一天，就有一群人会继续在这边，专专心心地做科研创新。”

如今研究院的愿景落在了醉心于科研的技术信仰者身上。说来宿命的是，就像某种轮回，当人工智能的概念挥别往日的“扭捏”而摇身蜕变为技术界的狂欢，这里的研究员或许比十八年前的前辈更为幸运。他们赶上了人工智能的黄金年代，可以肆意挥洒禀赋，紧跟这一技术的变迁孕育更多可能，且已收获颇丰。

随意摘取几例。2015年底，在ImageNet计算机视觉识别挑战赛——这一行业标杆属性的赛事上，微软亚洲研究院视觉计算组凭借深层神经网络技术的突破（他们使用了一种深度高达百层的神经网络，这要比过往任何成功使用的神经网络层数多5倍以上），获得图像分类，图像定位以及图像检测全部三个主要项目的冠军。几乎同一时刻，他们在另一项图像识别挑战赛MS COCO中的图像检测和图像分割项目上同样登顶——而一年过去，2016年10月，他们在MS COCO图像分割挑战赛中获得第一名，成绩比第二名高出11%，且相较去年第一名的成绩也有飞跃进步。

在用户一端，承载技术的是应用——这也是微软亚洲研究院近些年最令人欣喜的地方。

2015年，在中国登陆的Skype Translator实时语音翻译技术，就凝聚了研究院在语音识别，自动翻译和机器学习等多领域的成果，让世界上使用人数最多的语言和使用最广泛的语言直接对话成为可能。

而在风靡全球的How-Old.net（识别图片中人物的年龄）、Caption.ai(智能识图网站)等应用，同样基于研究院深耕多年的人脸识别技术。

当然，谈及人工智能应用，刷屏无数的小冰是绕不开的名字。抛开插科打诨不谈，我个人更看重的是它与其他行业的嫁接空间。譬如今年9月，小冰就进驻了“敦煌研究院”微信账号，摇身变为“敦煌小冰”，言谈之间告知人们关于敦煌的一切知识。嫁接得以实现，一项关键技术是微软亚洲研究院推出的自主知识学习技术（Doc Chat）：一种基于检索与排序直接从非结构化文档中选取句子作为聊天机器人回复的方法。这一技术也已运用于2016年8月发布的第四代微软小冰跨平台商业解决方案，可让小冰更轻松地学习不同领域知识，完成更多嫁接可能。

更多底层基础突破和应用在此不表，微软亚洲研究院——以及整个微软对人工智能的宏愿，都如现已负责微软人工智能与微软研究院事业部的沈向洋所言：“我们专注于创造一个人工智能集群，它不仅覆盖基础设施、服务、应用以及智能助理，而且可为包括消费者、企业用户和开发者在内的众多客户群所用。我们正在创建许多工具，让繁忙的职场人士在工作上能够有条不紊；让身处不同国家的亲朋好友能够跨越语言障碍尽情交谈；让智能手机用户能够分身有术、更快捷地发送信息。同时，我们也在为企业用户提供其所需的工具，让他们能够将智能注入到每一个产品和每一项商业决策中去。”

总之，尽管换了“标题”和讲述者，但这里与人工智能从十八年前开始诉说的故事，仍在继续，而这本书籍也因技术的积淀而变得更为厚重。

● 基因传承

故事得以延续的前提是传承，研究为先，自由开放的环境与体系得以让这里相对纯粹的“传帮接带”。而这一切都在每位院友的记忆中留下了深深地烙印，并正在各自领域里将当年的学习和感悟不断“释放”。

事实上，那些离开的人几乎无一例外地珍视这段纯粹的职业生涯。而当他们追忆微软岁月，“感谢”与“传承”总是并列出现。“在现在的创业路上，研究院的文化氛围也深深的影响着我们的公司文化，感谢研究院，感谢当年指导我们的Mentors，感谢一起实习的小伙伴们。”已在创业的蔡东翔院友这样说道，“希望Mentor文化可以在院友会得到一个很好的传承。”

同样作为院友，清华大学计算机系副教授朱军则将传承细化：“我在研究院实习3年多……现在还清晰记得，第一天面试时问我‘为何读博士？’，这成了我现在面试学生必问的题目。Harry（沈向洋）的‘work hard, play harder’是我记得最清楚的一句话，现在也拿来教育学生。”

传承仍在继续。在不少院友看来，如今，院友会的成立可以发挥桥梁与纽带作用，分享经验与资源，甚至在某些奇特的瞬间感到从未离开过一样。这有点像是管理界颇为盛行的联盟关系，当人们离开一个地方，可以通过“同事联络网”成为这个地方的支持者和外部资源，并将它身上的基因散播到更远处。

好吧，也许正如罗振宇所说：世界上至少有两种游戏，一种是有限游戏，一种是无限游戏。有限游戏以游戏结束为目的，比的是谁能赢；无限游戏则是让游戏本身继续下去，而不是让谁赢，卷入的资源越多，参与的人数越多，游戏就能继续下去，这才是无限游戏的目标，也正是那些“老公司”的价值所在。

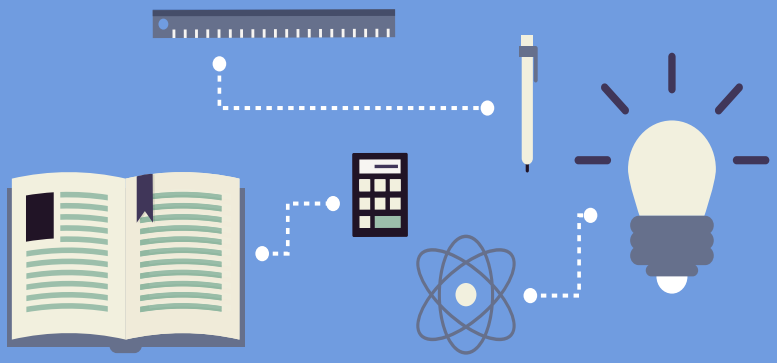
在我看来，这也正是微软亚洲研究院——这一席“流动的盛宴”，留给技术世界的真正财富。

第十八届
“二十一世纪的计算大会”

2016年11月，由微软亚洲研究院举办，以"Human and Machine，Working as a Team"为主题的第18届"二十一世纪的计算"大型国际学术研讨会在韩国首尔成功举办。众多计算机领域顶级大师在此次大会上分享了各自独特且深远的见解，并深入探讨在人工智能与人类智慧的结合下，改变社会的无限潜能。

扫描二维码
观看演讲视频

知无不言，言无不尽
TO BE CONTINUED



微软认知服务：人工智能人人皆享

在推进AI普及化进程中，微软认知服务扮演着重要角色。作为25款工具集合，微软认知服务集微软机器学习和AI研究之大成，免去广大开发人员从无到有的自创技术之苦，让毫无机器学习背景的开发人员也可在其应用中轻松添加诸多智能功能。近日，微软发布了认知服务工具的公开预览版，“内容审阅”和“必应语音API”也已正式上线。本文带你解锁更多新故事和新体验。

李飞飞、李凯、沈向洋、张钹、洪小文：我们正处于人工智能的秋天

未来论坛2017年会可谓大咖云集，星光闪耀。在微软全球资深副总裁、微软亚太研发集团主席兼微软亚洲研究院院长洪小文主持的未来智能讨论环节，斯坦福终身教授李飞飞、美国工程院院士李凯、微软全球执行副总裁，微软人工智能及微软研究事业部负责人沈向洋、清华大学教授张钹院士四位AI泰斗级人物共谈AI未来。华人AI大牛们表示：我们正处于人工智能的秋天。

微软亚洲研究院开源图数据查询语言LIKQ

近日，微软亚洲研究院在GitHub上开源图数据查询语言LIKQ。它是基于分布式大规模图数据处理引擎Graph Engine的一种可用于子图和路径查询的数据查询语言，可以让开发人员无需学习新的领域相关的特定查询语言，直接使用原生C#代码即可构建知识图谱语言，从而使海量图数据的实时检索和集成变得触手可得。

张冬梅：带领数据探索的梦之队

13年前，刚回国的张冬梅对微软亚洲研究院的生活既有期许，也有压力。13年后，张冬梅已是首席研究员，带领的软件分析组也成绩斐然。他们通过与微软Power BI产品组的合作，将“快速洞察”、“信息图设计”等技术成功地转化到Power BI产品中。同时，两个团队下一步的合作更是为 Power BI 构建了良好的产品愿景，让该产品在商业智能领域最权威的报告之一——Gartner魔力象限2017中身居领导者象限的高位，并在未来布局能力上一骑绝尘，领先所有厂商。本文带你走进这支数据探索梦之队背后的故事。

微软亚洲研究院

微软亚洲研究院是微软公司在亚太地区设立的研究机构，也是微软在美国本土以外规模最大的一个。从1998年建院至今，通过从世界各地吸纳而来的专家学者们的鼎力合作，微软亚洲研究院已经发展成为世界一流的计算机基础及应用研究机构，致力于推动整个计算机科学领域的前沿技术发展，并将最新研究成果快速转化到微软全球及中国本地的关键产品中，帮助消费者改善计算体验。同时，微软亚洲研究院着眼于下一代革命性技术的研究，助力公司实现长远发展战略和对未来计算的美好构想。

众多从微软亚洲研究院诞生的创新技术转移到了微软的产品中，包括：Office、Windows、Azure、Bing、Xbox Kinect等，以及近年来以微软小冰、Cortana、Microsoft Translator、微软认知服务等为代表的人工智能产品。同时，基于以实践驱动的研究理念，近年来微软亚洲研究院还孵化了很多广受欢迎的应用和技术项目，如微软小英、微软识花、微软对联/字谜、Urban Air等等。

微软亚洲研究院一直努力营造奋发、进取的科研环境，倡导对技术进步怀有远大抱负，推崇大胆创新以及富于冒险的极客创新精神。研究院鼓励研究人员树立长远眼光，加强与外界的交流，在第一时间接触世界领先的研究成果，对行业发展、技术趋势、及消费者需求高度敏感、迅速反应。微软亚洲研究院强调成员之间的相互信赖、相互尊重与开放合作，并承诺与高校和科研机构开展持久而有效的合作，促进协作、激发创新、推进教育，变梦想为现实！

4,000+
论文发表

IMAGENET
COCO
Common Objects in Context

教育部最佳合作伙伴

最火的计算机实验室

最佳企业研究院

优秀博士后科研工作站

50+
最佳论文

微软在美国本土以外规模最大的研究机构

5 大研究方向

Microsoft

微信

知乎

电话：86-10-59178888

传真：86-10-82868527

网址：http://www.msra.cn/

微博：http://t.sina.com.cn/msra